LA-UR-14-23875
Approved for public release; distribution is unlimited.

Title: Mama Software Features: Uncertainty Testing

Author(s): Ruggiero, Christy E.
Porter, Reid B.

Intended for: Web

Issued: 2014-05-30

MAMA Software Features:
Uncertainty Testing

Abstract:

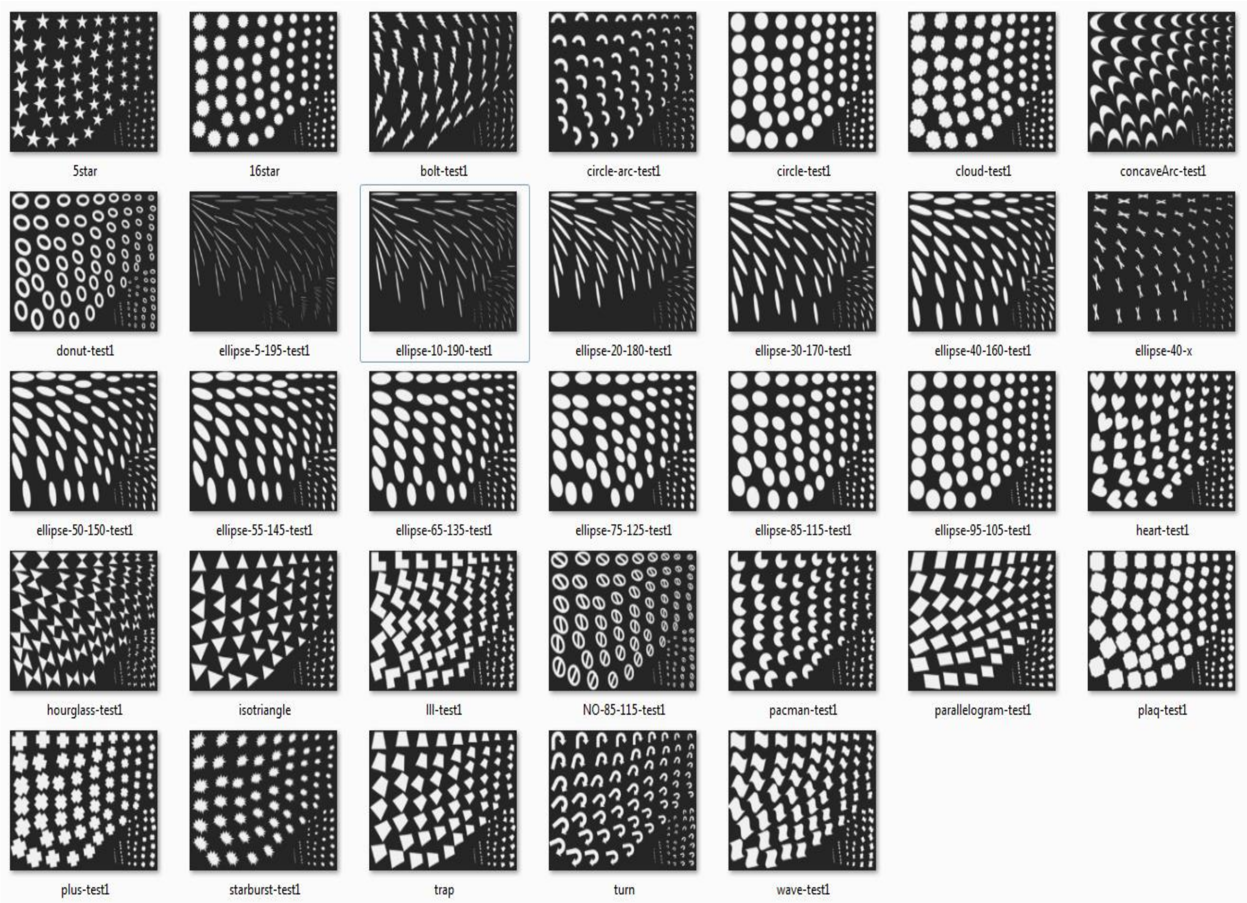
This document reviews how the uncertainty in the calculations is being determined with test image data. The results of this testing give an 'initial uncertainty' number than can be used to estimate the 'back end' uncertainty in digital image quantification in images. Statisticians are refining these numbers as part of a UQ effort..

Contents

1.0 Overview:

We have continued to verify the software calculations and determine the uncertainty in the calculations using synthetic images. As described in previous reports, synthetic images are the most efficient method for verifying the functionality and correctness of the calculations and establishing the operational space and uncertainty in the calculations. The results of these tests are undergoing analysis by a statistician (Tom Burr) in order to formally address reproducibility, uncertainty and bias. Until then, we consider the percent standard deviation that we calculate herein a number that can be used as an 'initial uncertainty' estimate in order to give users a sense of the magnitude of the innate uncertainty in digital image quantification.

The predominate set we are using for testing 33 shapes each at 15 sizes X 7 rotations for the size and shape attribute calculations. These shapes were generated in drawing software, and are 'grayscale' images, so are not simply white shapes on a black background, but have a gradient edge that would be more typical of real particle images. This edge allows for more variation in the segmentation than would a hard black/white edge, which will give variations more typical of real images.



. .

2.0 Calculation Changes:

Through the rounds of testing and verifying the calculations, we corrected the following calculations and then reassessed the calculation variations:

1. Vectorization was corrected for Bias
2. Ellipse aspect ratio and all ellipse-based calculations were corrected for errors that results when the best fit ellipse by the calculation method was not a best representative ellipse. This also corrected cases where the calculation failed.
3. Chordal aspect ratio was corrected for large variance based on a 5 degree tolerance on the vectorized perimeter. The correction is fundamentally similar to the ellipse correction. This removed one source of variation in these calculations.

Not yet corrected: The calculation settings on the chordal aspect ratio do not take into account multiple near-equivalent maximum chordal diameters, leading to large variation in these calculations with some convex shapes for which many maximum chordal diameters can be drawn. We are in the process of fixing this calculation. The initial fix was computationally too expensive (i.e. too slow), so is being optimized and tested. Therefore, in the results below, the standard deviation in the chordal aspect ratio is LARGER than it will be once the improved calculations are fully implemented.

3.0 The Test Results:

Presented below are selected uncertainty summary plots that were generated from analyzing the size-shape-rotation data set. Each image was imported into the MAMA software and segmented. The segments were then refined by hand to include any missed object that did not segment (often an issue with the smallest size sets, which are sometimes only a few pixel in area.). For calculations that should be nominally invariant across size, the data average and standard deviation was calculated in size bins (0-10; 11-20; 21-30; 31-40; 41-60; >60 Pixel ECD) and plotted. The shapes were separated into 3 categories: those with aspect ratio less than 5, those with aspect ratio 5-10, and those with aspect ratio > 10. The % standard deviation for all shapes in each shape category was then determined. This % standard deviation can be used as an estimate of the pixel-based uncertainty for object that fall into the bin/category. (This 'initial uncertainty' is being analyzed and refined by statisticians.) The high aspect ratio data will need further analysis since we had only a few shapes at these aspect ratios, and none with convex or irregular edges. These high aspect ratio tests will necessarily be limited to only large size ranges. The variation in this data reflect predominately the pixel-based representation and math uncertainty, and reflect some of the uncertainty in object segmentation. Segmentation routine variability and errors has also been tested with this data set, but data analysis for that work is not yet complete.

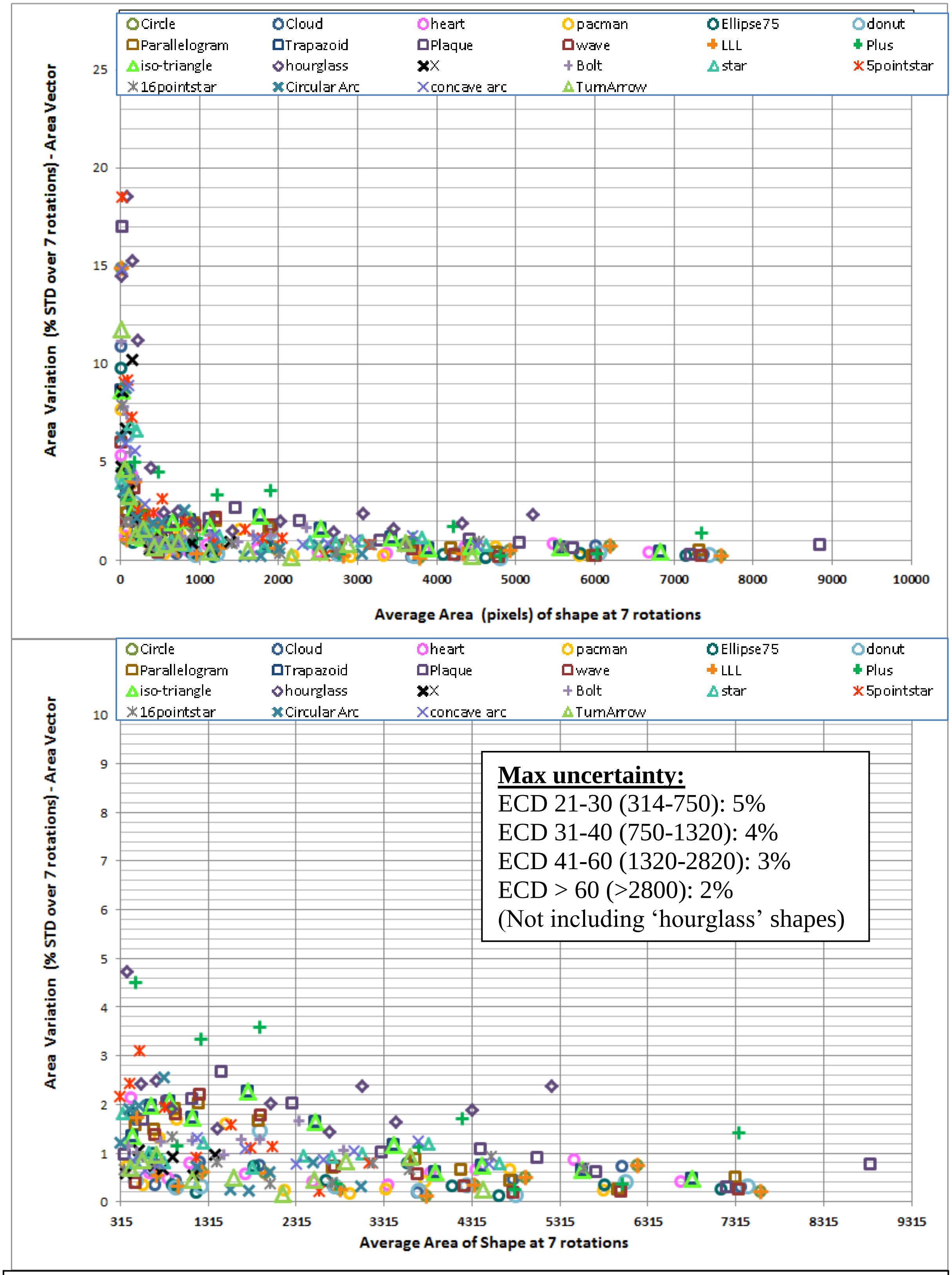
We will further refine this data using a third attribute for uncertainty 'bins' (Size, Aspect Ratio, Area convexity or perimeter convexity) because the convexity increases the variability in many of the attributes, while the variability in compact,, non-convex shapes is lower. The numbers reported on the graphs below only represent size and aspect ratio bins.

Initial tests we conducted suggested that to achieve a calculation uncertainty in area (pixel vs. vector area) of 10% or less, particles need to be at least 20 pixels in equivalent circle diameter

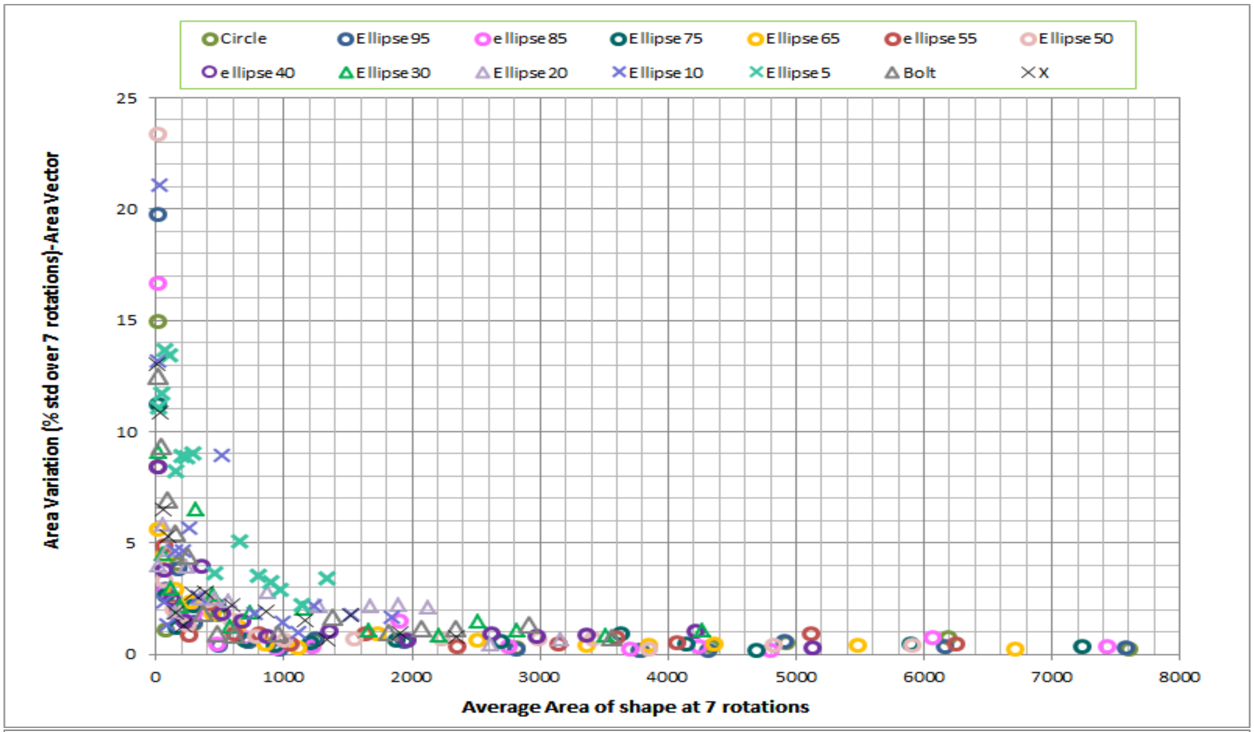
(ECD) for most regular shapes, but may need to be slightly larger (25 pixels) for very convex or elongated shapes. The uncertainty determined in the calculations using this data sets also agrees with the **20-25 pixel ECD limit for <10% variability in calculations.**.

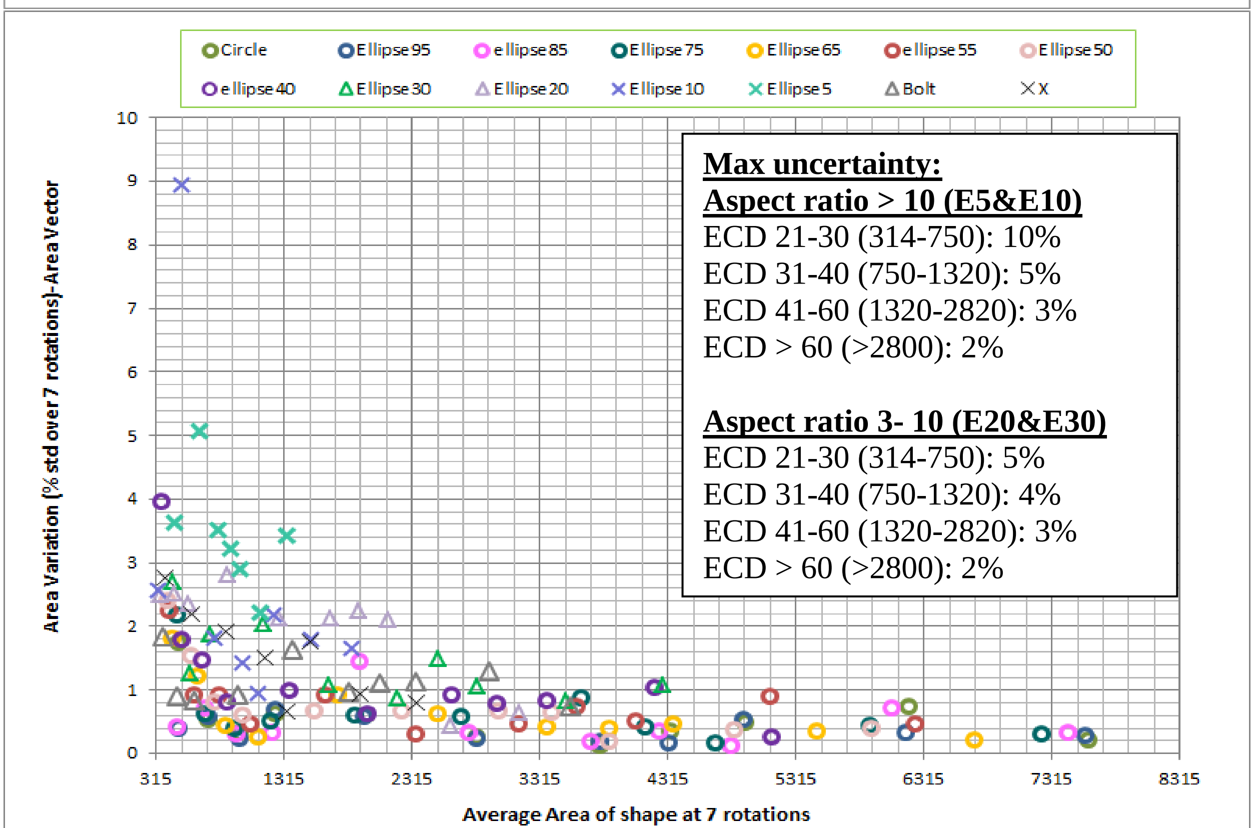
These tests also show, as would be expected, that any features on an object need to be at a reasonable size in order to quantify. For examples, errors become >10% on shapes that are very narrow (high aspect ratio ellipses with a width of < 5 pixels), have narrow features (16 point stars with 'rays' that are only a few pixels in width). **Shapes that have near inversion centers** of only a few pixels (hourglass shapes) were already listed in the attribute calculations as being problematic (based on the openCV library code warnings), and are shown in the plots below to show the error-in the calculations. These issues will need to be added as additional warnings to users for acquiring images in which the objects have enough pixel coverage to be accurately quantified. Nominally, we would state that **no feature of interest should be less than 5 pixels** to insure accurate shape quantification.

Related to that limit, attributes such as Circularity, Roundness, Area Convexity, and Perimeter Convexity will actually change with object size due to the limitation of the pixel resolution. The "20 pixel" ECD should eliminate most issues, but in extremely convex shapes or shapes with detailed features, caution should be used before size-invariance is assumed.

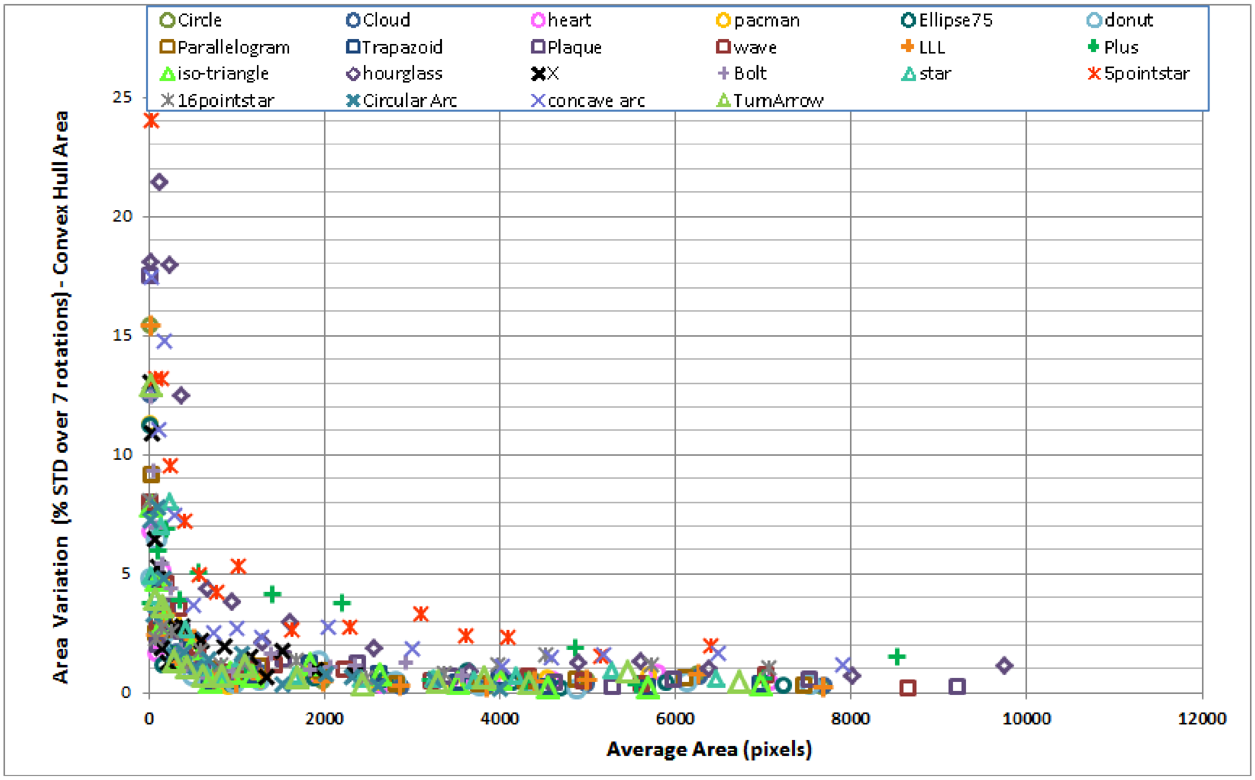


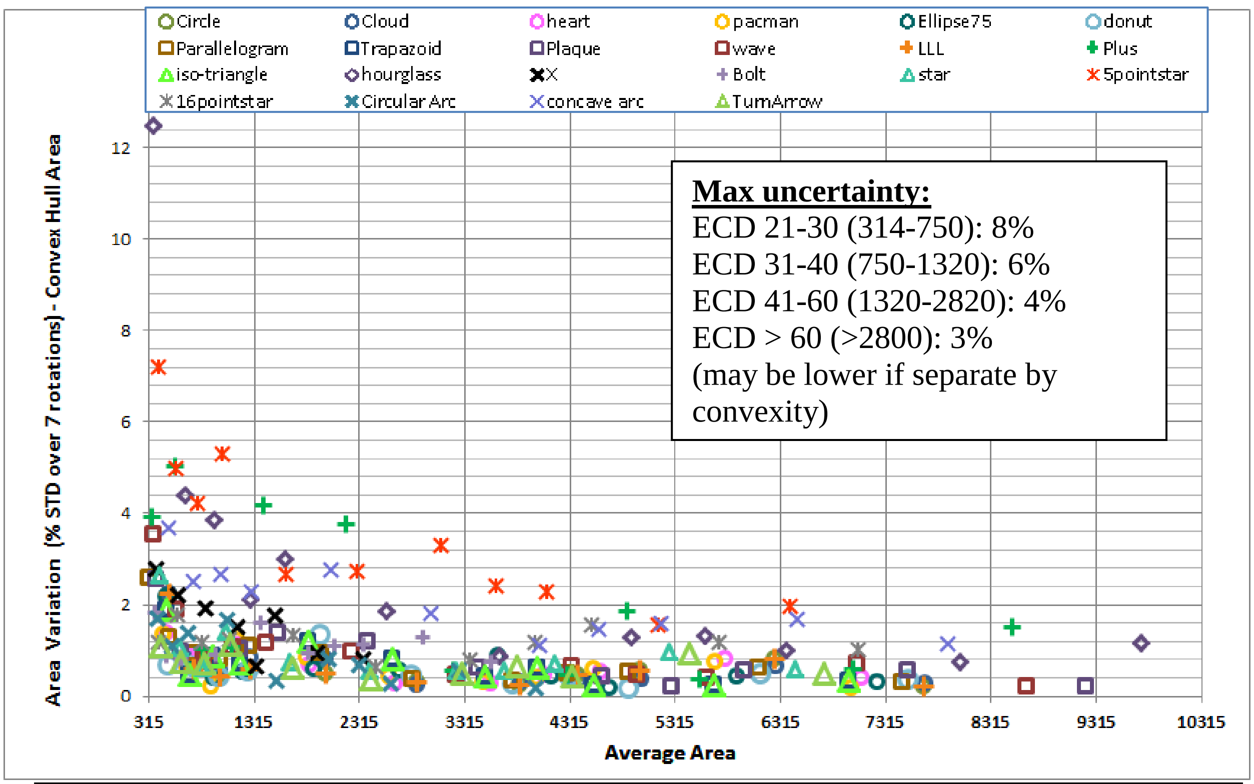
Vector Area for Aspect ratio < 3(in pixels) plotted against the area variation (% STD) over the 7 rotations at each size. The Top figure shows the full data, the bottom figure shows the region for ECD > 20 (Area >314 pixels)



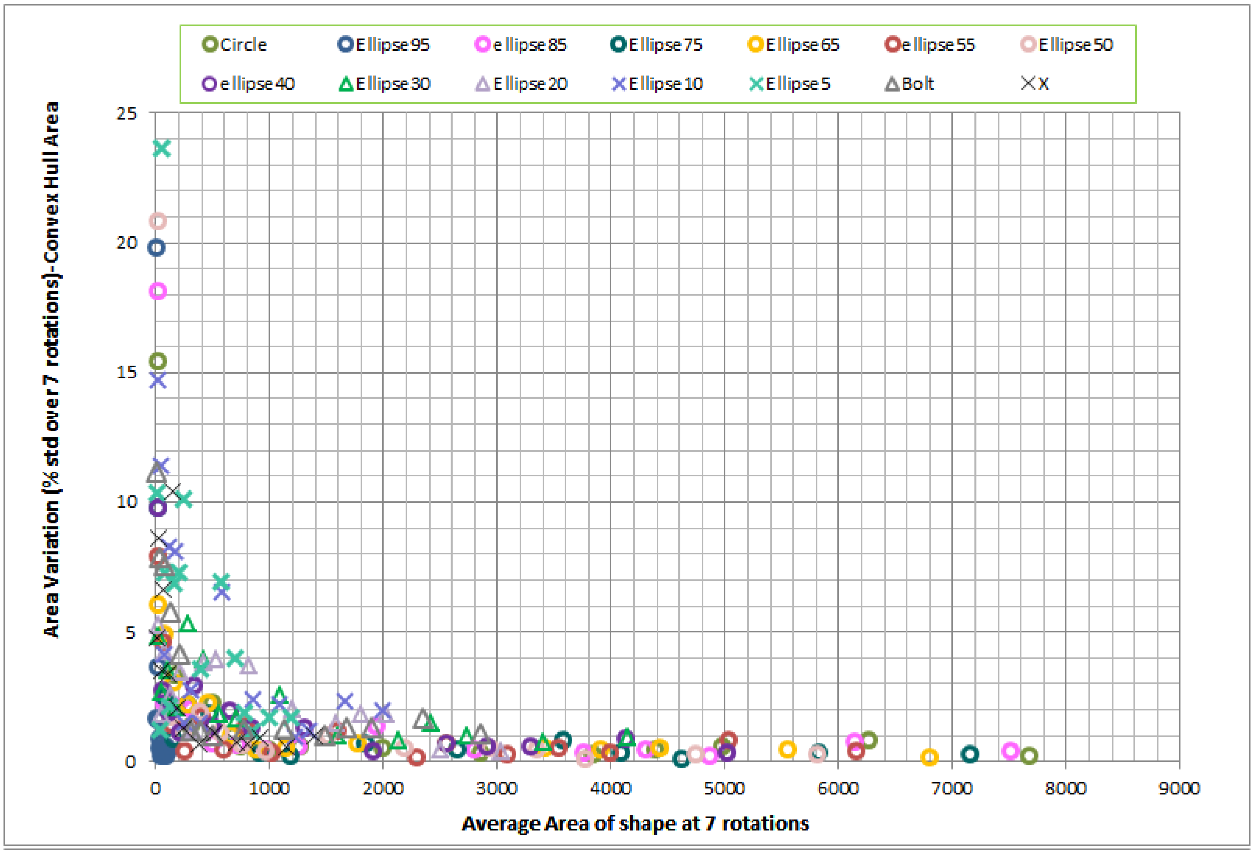


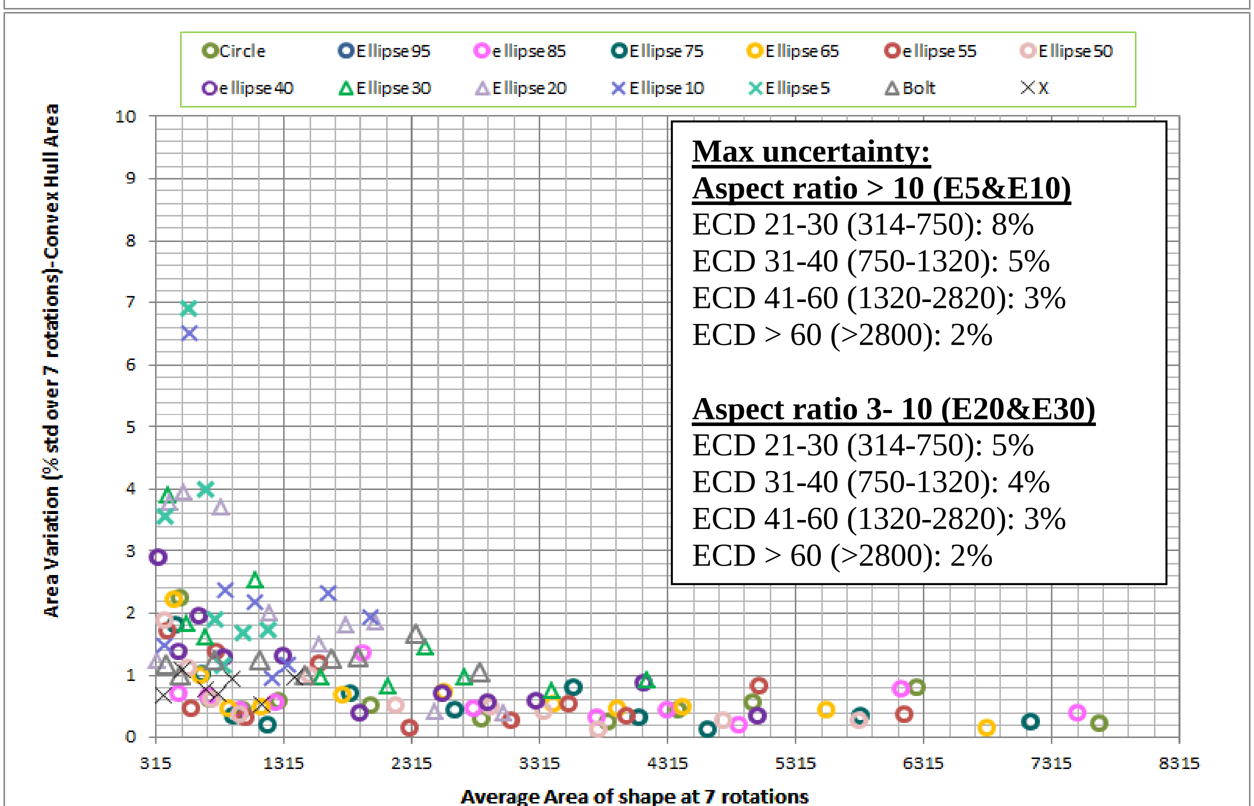
Vector Area for ellipses with range of Aspect Ratios (in pixels) plotted against the area variation (% STD) over the 7 rotations at each size. The Top figure shows the full data, the bottom figure shows the region for ECD > 30 (Area >314 pixels). These shapes range from aspect ratio 1 to >20.



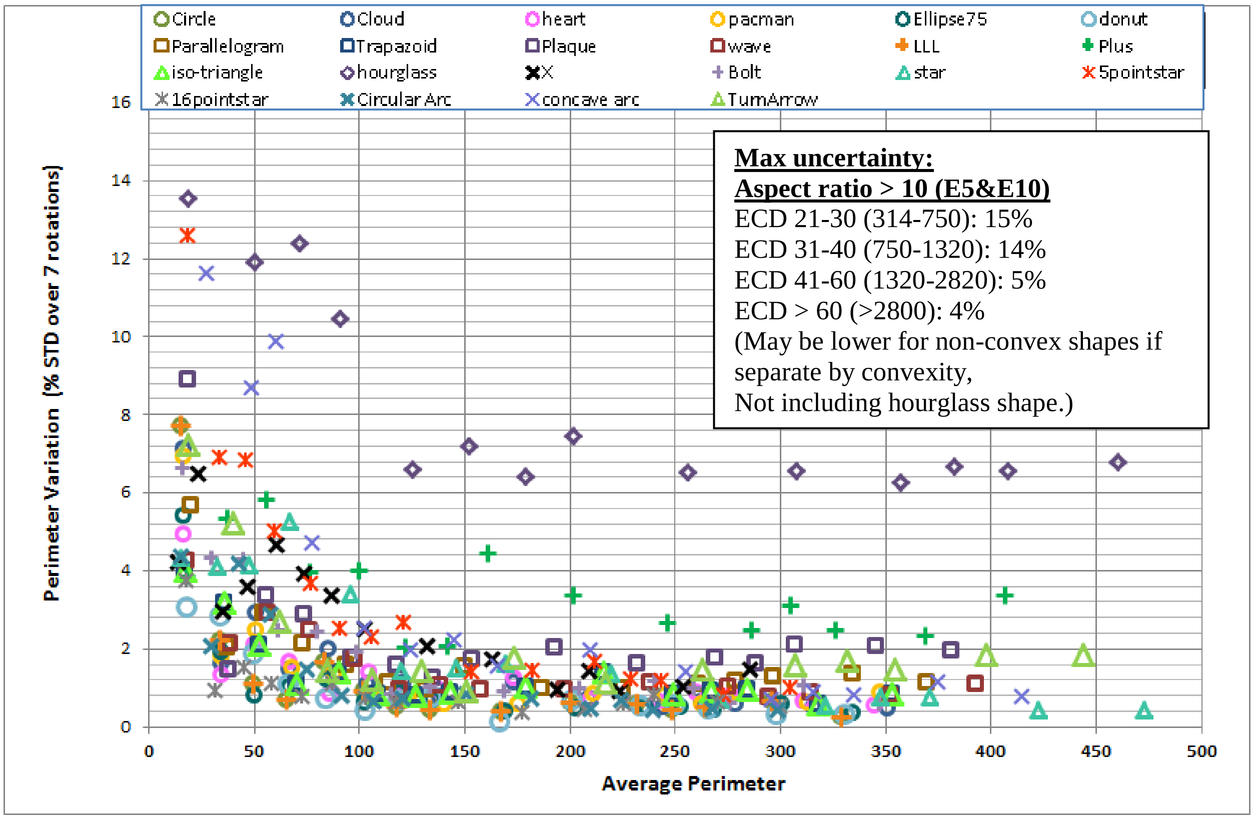


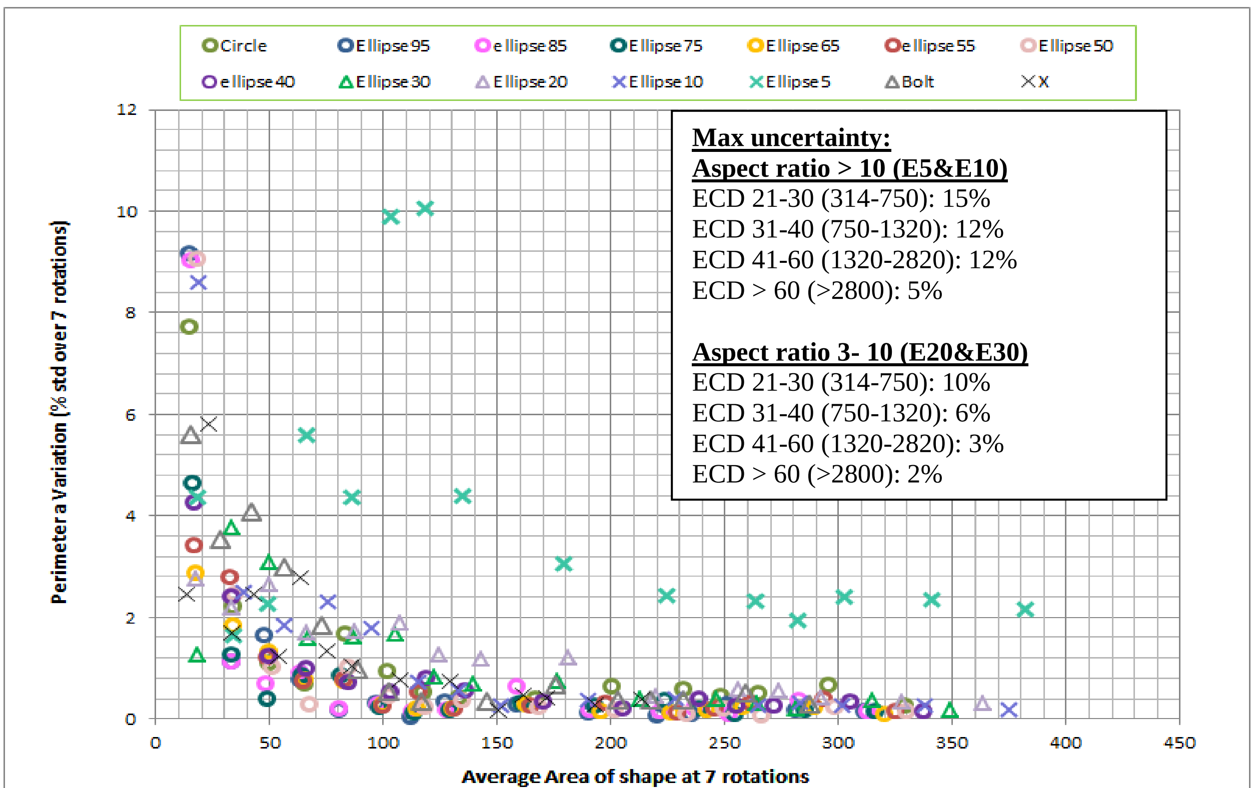
Vector Convex Hull Area for Aspect ratio < 3 (in pixels) plotted against the area variation (% STD) over the 7 rotations at each size. The Top figure shows the full data, the bottom figure shows the region for ECD > 30 (Area >314 pixels).



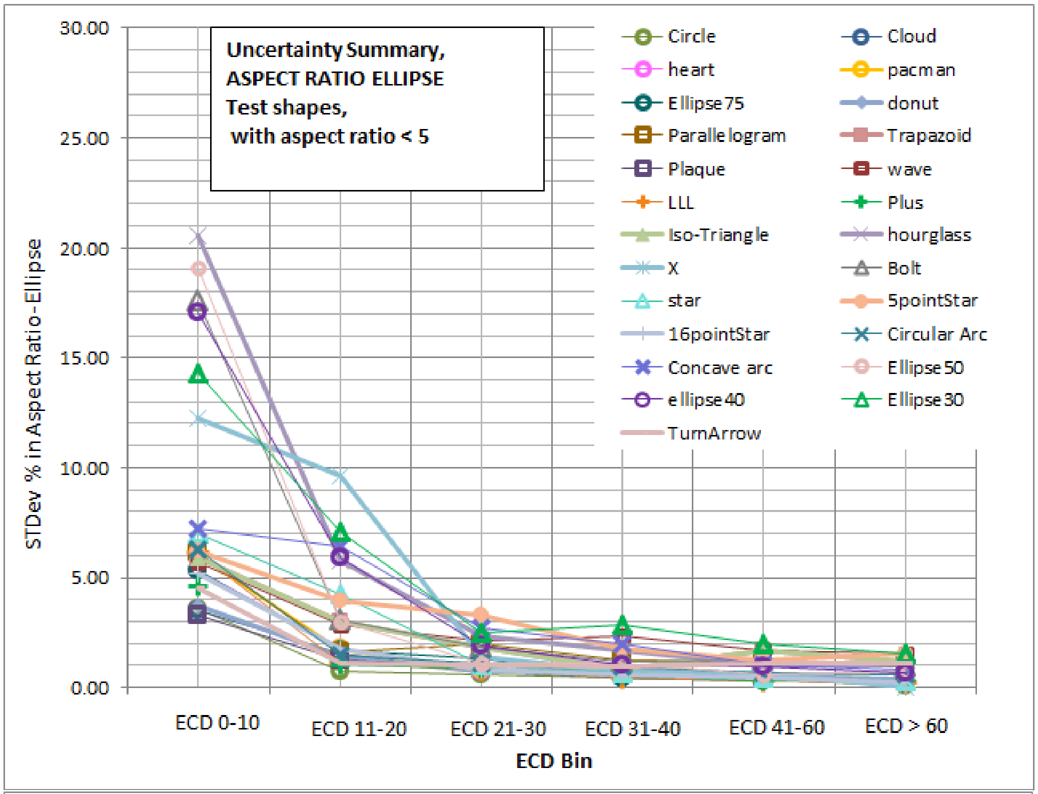


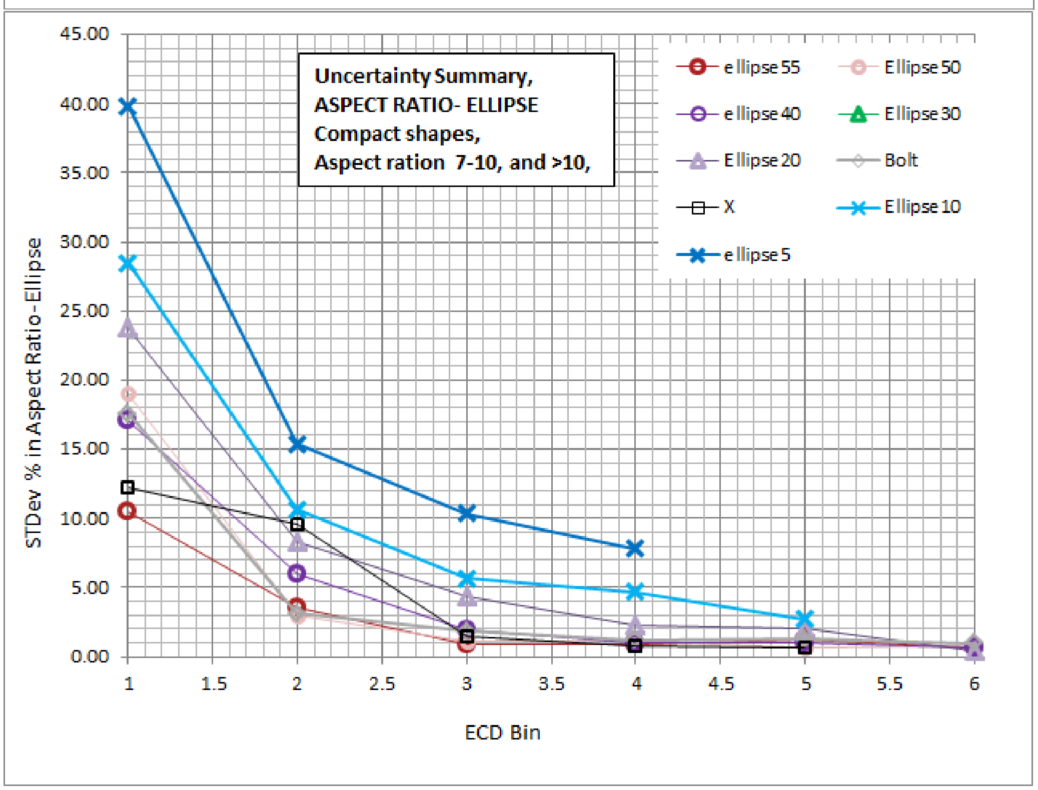
Vector Convex Hull Area for ellipses with range of Aspect (in pixels) plotted against the area variation (% STD) over the 7 rotations at each size. The Top figure shows the full data, the bottom figure shows the region for ECD > 30 (Area >314 pixels).





Vector Perimeter (in pixels) plotted against the area variation (% STD) over the 7 rotations at each size. The Top figure shows a variety of shapes with aspect ratio <3; the bottom figure shows the same plot for a series of ellipses with a range of aspect ratios.

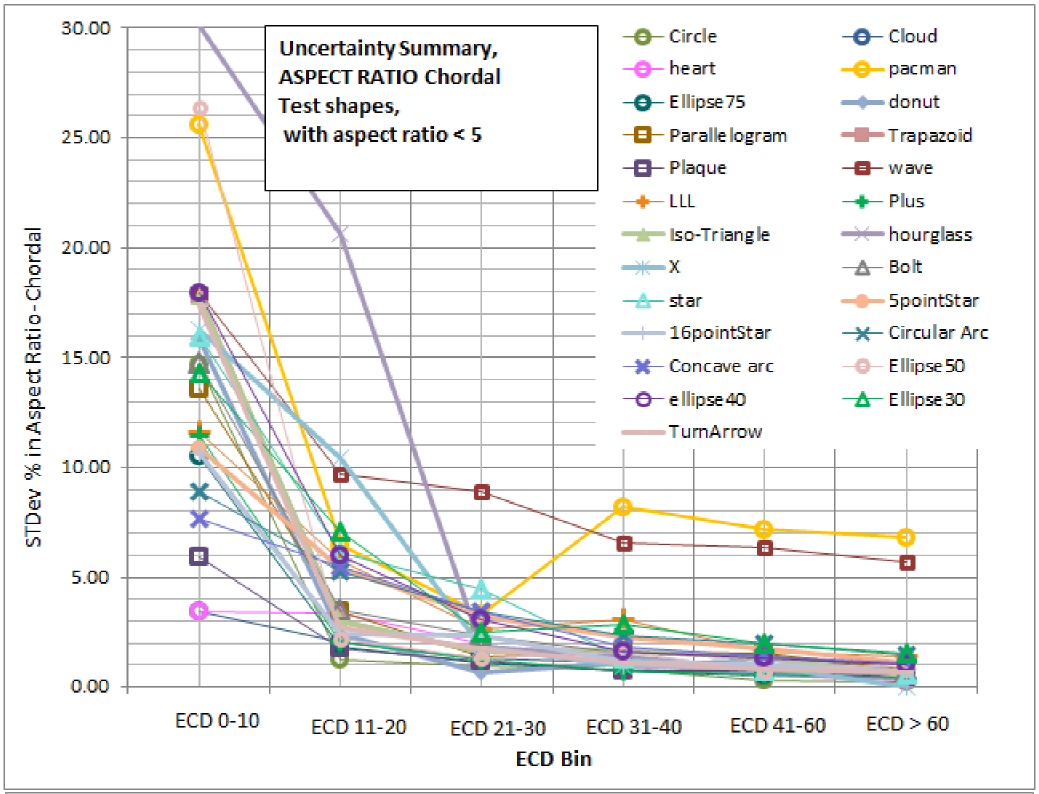


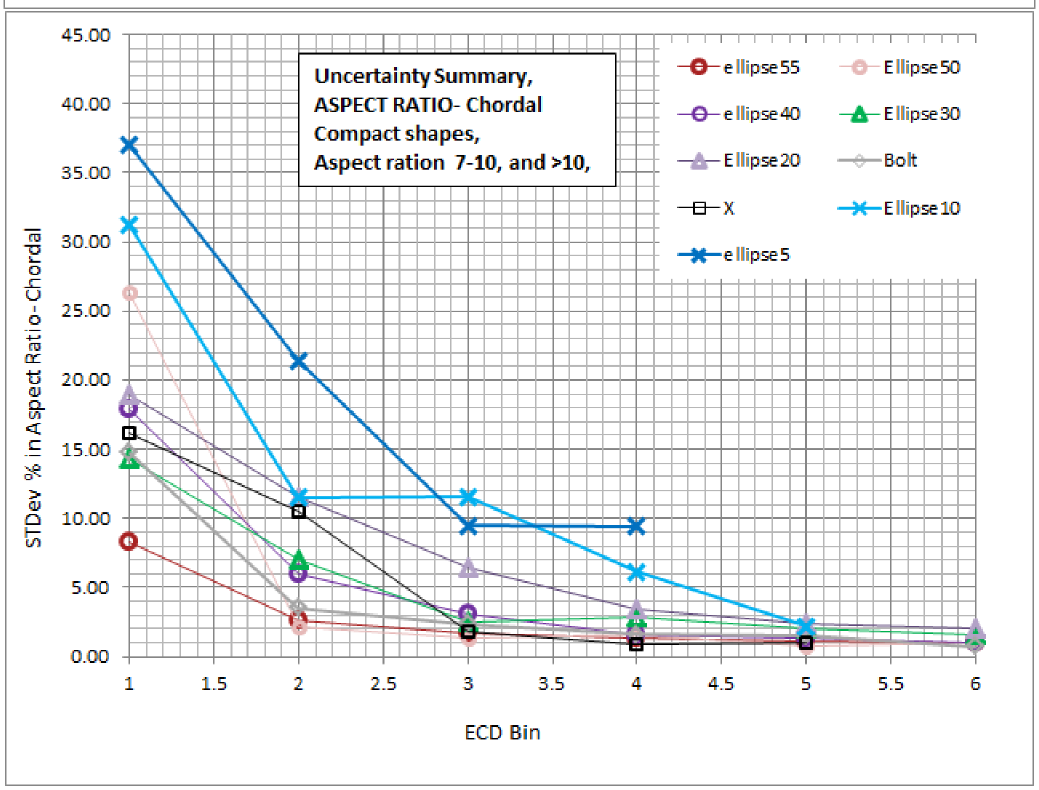


Top: Calculated Uncertainty in Aspect Ratio Ellipse for shapes with aspect ratio < 5. The plot shows the % Standard Deviation (Std Dev/ mean) calculated for each shape independently (25 shapes) plotted by size bin.

The maximum % STD for each bin (top data point) can be used as an initial uncertainty estimate, until full statistical analysis of the data is completed. .

Bottom: Same plot as above, but for shapes with aspect ratio > 10 (Ellipse 5 & 10) and Aspect ratio 3-10 (Ellipse 20 & 30).



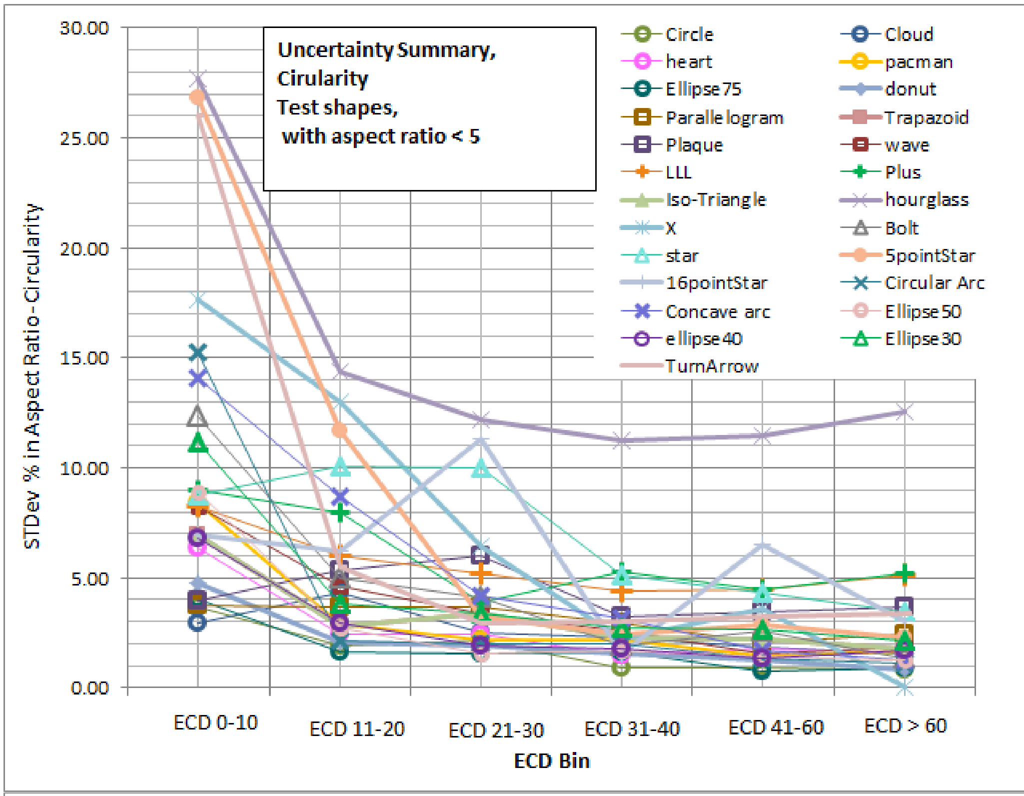


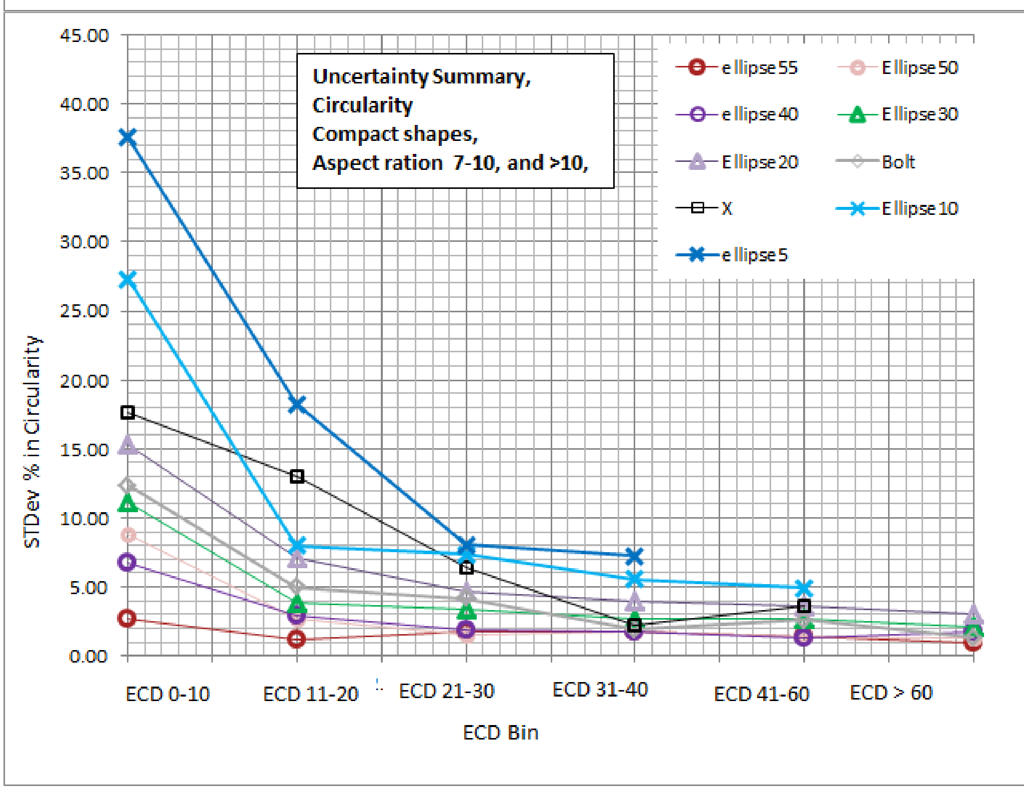
Top: **Calculated Uncertainty in Aspect Ratio – Chordal** for shapes with aspect ratio < 5. The plot shows the % Standard Deviation (Std Dev/ mean) calculated for each shape independently (25 shapes) that were in each size bin.

The maximum % STD for each bin (top data point) can be used as an initial uncertainty estimate, until full statistical analysis of the data is completed. .

The uncertainty for this attribute will significantly reduce when the calculation is updated, as described in previous reports.

Bottom: Same plot as above, but for shapes with aspect ratio > 10 (Ellipse 5 & 10) and Aspect ratio 3-10 (Ellipse 20 & 30).





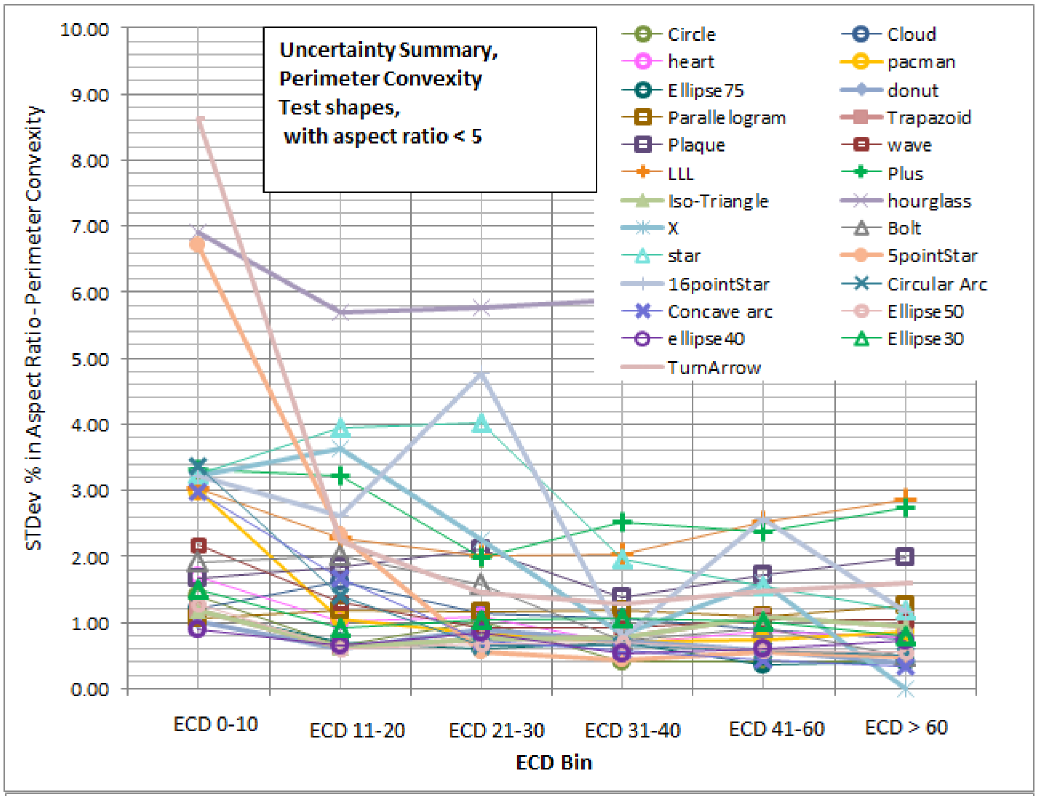
Top: Calculated Uncertainty in Circularity for shapes with aspect ratio < 5. The plot shows the % Standard Deviation (Std Dev/ mean) calculated for each shape independently (25 shapes) that were in each size bin.

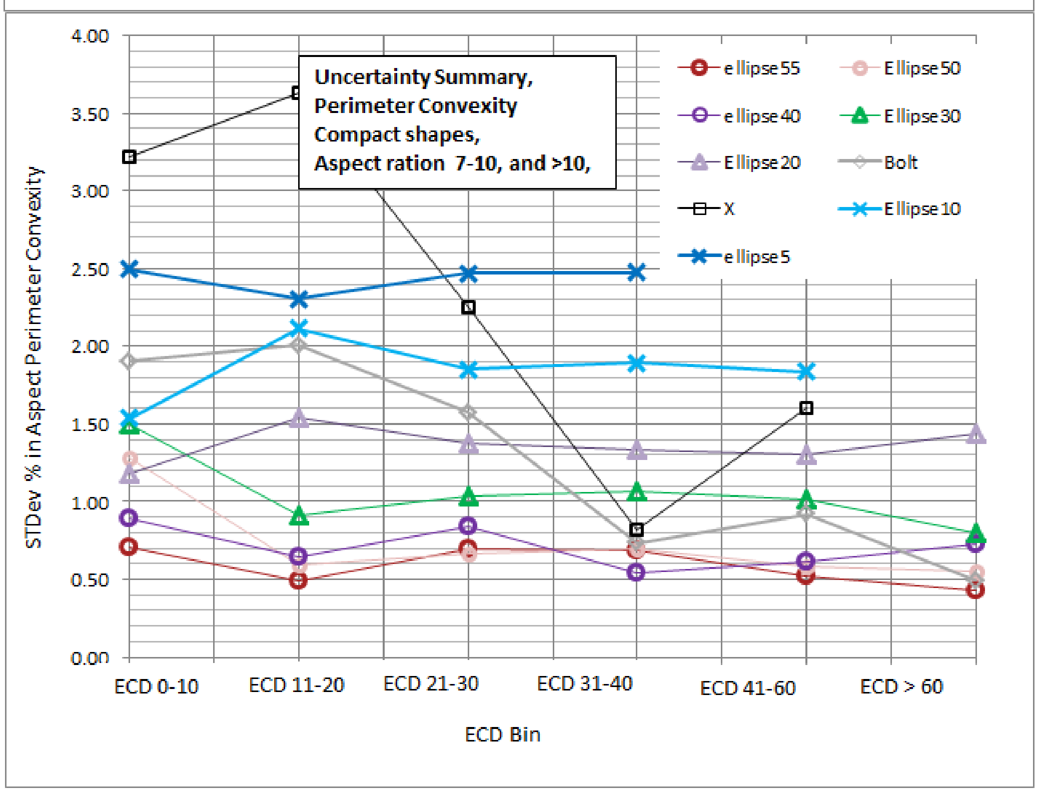
The maximum % STD for each bin (top data point) can be used as an initial uncertainty estimate, until full statistical analysis of the data is completed. .

The uncertainty for this attribute may reduce if we separate by convexity as well as size/aspect ratio, since the most convex shapes have the largest % STD.

(Note: Hourglass shape calculation issue can be readily seen here.)

Bottom: Same plot as above, but for shapes with aspect ratio > 10 (Ellipse 5 & 10) and Aspect ratio 3-10 (Ellipse 20 & 30). These uncertainties need to be confirmed with a set of convex and non-ellipse shapes.





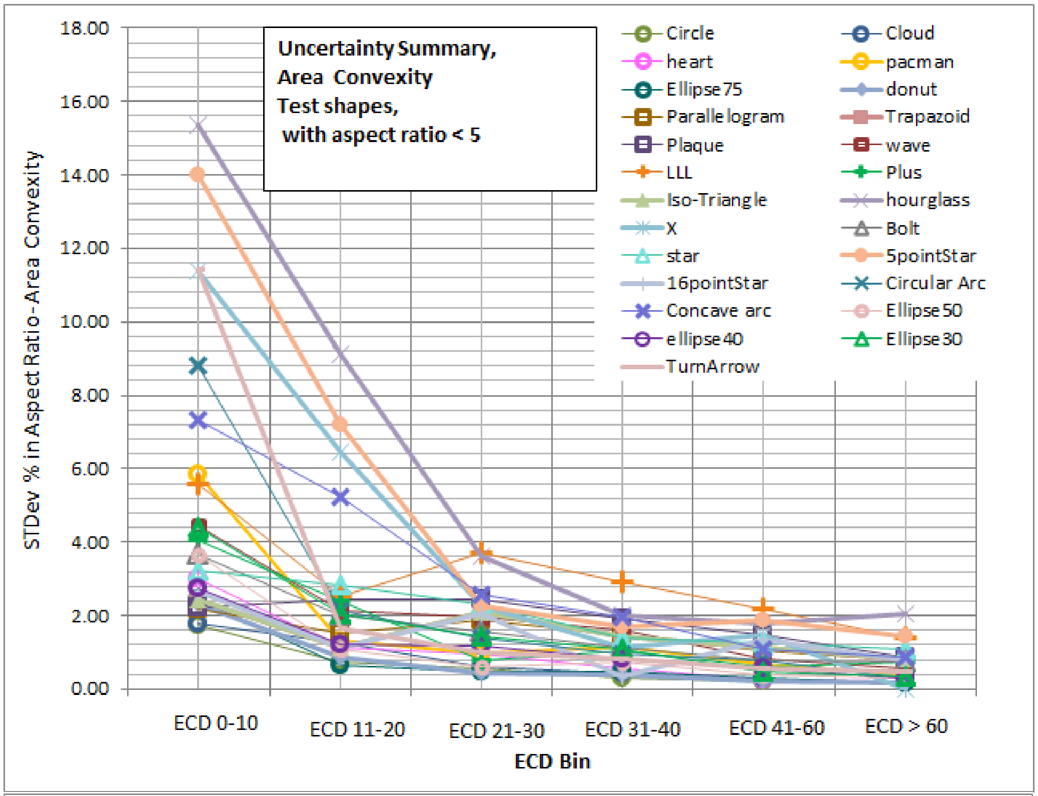
Top: **Calculated Uncertainty in Perimeter Convexity** for shapes with aspect ratio < 5. The plot shows the % Standard Deviation (Std Dev/ mean) calculated for each shape independently (25 shapes) that were in each size bin.

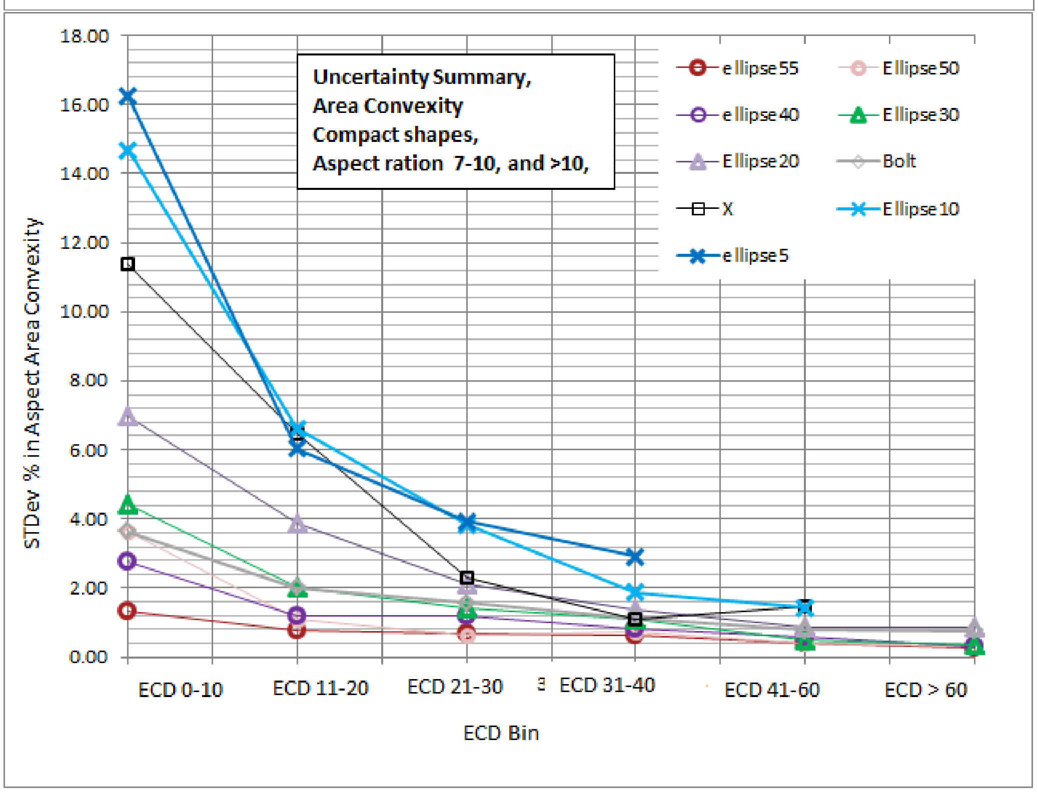
The maximum % STD for each bin (top data point) can be used as an initial uncertainty estimate, until full statistical analysis of the data is completed. .

The uncertainty for this attribute may reduce if we separate by convexity as well as size/aspect ratio, since the most convex shapes have the largest % STD.

(Note: Hourglass shape calculation issue can be readily seen here.)

Bottom: Same plot as above, but for shapes with aspect ratio > 10 (Ellipse 5 & 10) and Aspect ratio 3-10 (Ellipse 20 & 30). These uncertainties need to be confirmed with a set of convex and non-ellipse shapes.





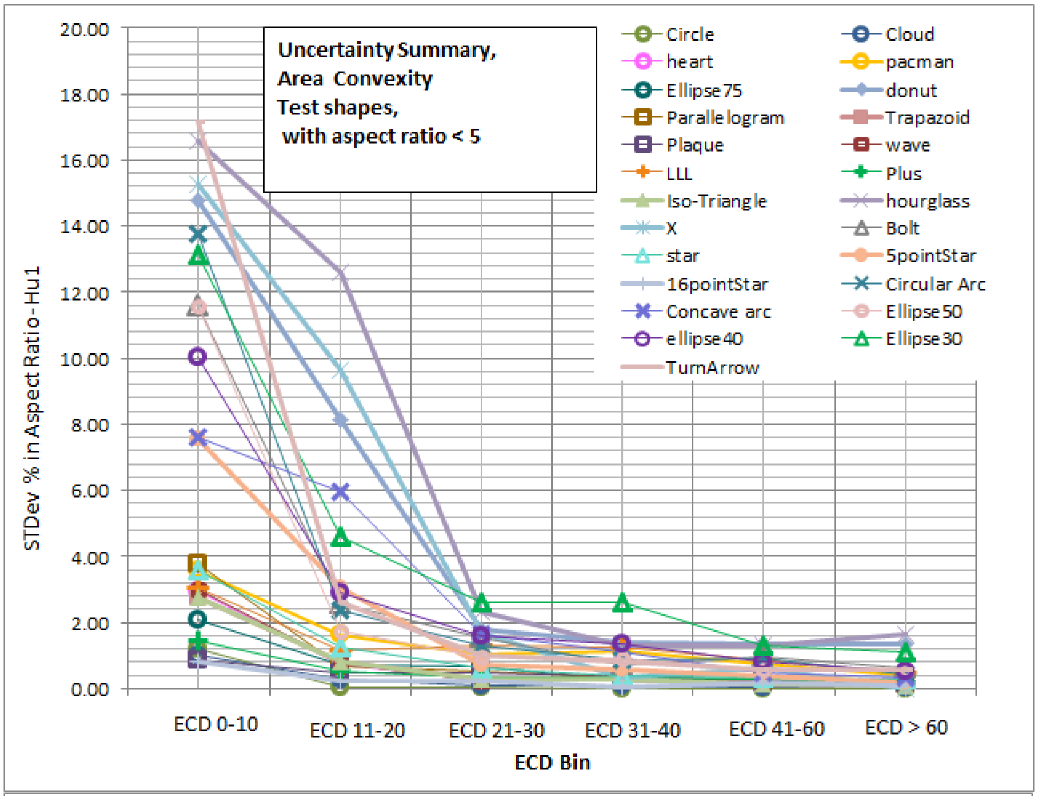
Top: Calculated Uncertainty in Area Convexity for shapes with aspect ratio < 5. The plot shows the % Standard Deviation (Std Dev/ mean) calculated for each shape independently (25 shapes) that were in each size bin.

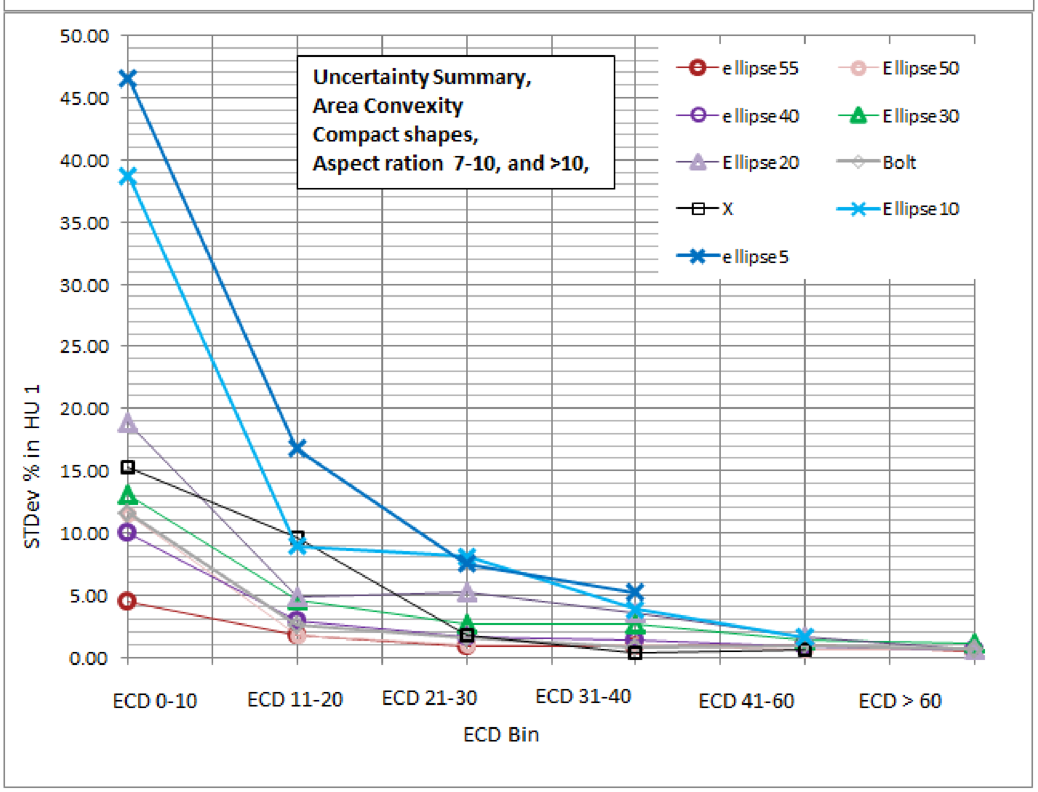
The maximum % STD for each bin (top data point) can be used as an initial uncertainty estimate, until full statistical analysis of the data is completed. .

The uncertainty for this attribute may reduce if we separate by convexity as well as size/aspect ratio, since the most convex shapes have the largest % STD.

(Note: Hourglass shape calculation issue can be readily seen here.)

Bottom: Same plot as above, but for shapes with aspect ratio > 10 (Ellipse 5 & 10) and Aspect ratio 3-10 (Ellipse 20 & 30). These uncertainties need to be confirmed with a set of convex and non-ellipse shapes.





Top: **Calculated Uncertainty in Hu-1** for shapes with aspect ratio < 5. The plot shows the % Standard Deviation (Std Dev/ mean) calculated for each shape independently (25 shapes) that were in each size bin.

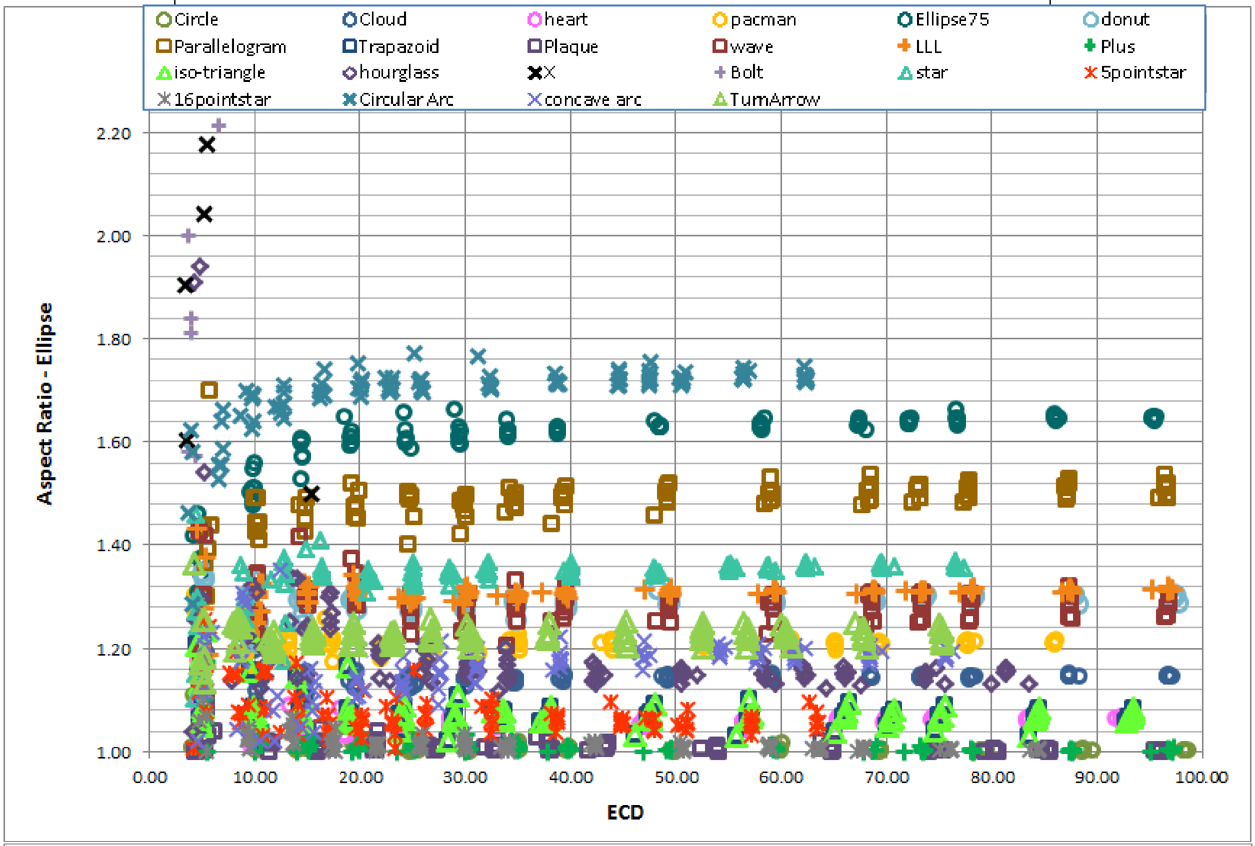
The maximum % STD for each bin (top data point) can be used as an initial uncertainty estimate, until full statistical analysis of the data is completed. .

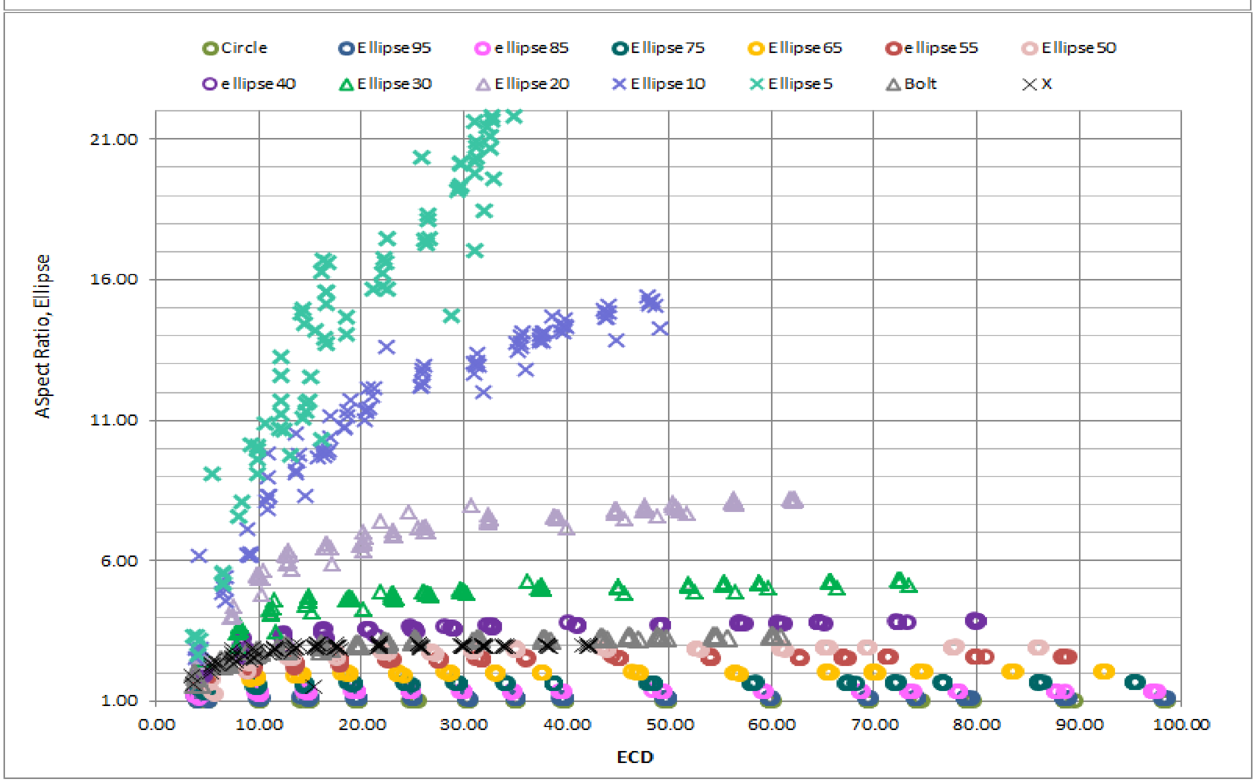
The uncertainty for this attribute may reduce if we separate by convexity as well as size/aspect ratio, since the most convex shapes have the largest % STD.

(Note: Hourglass shape calculation issue can be readily seen here.)

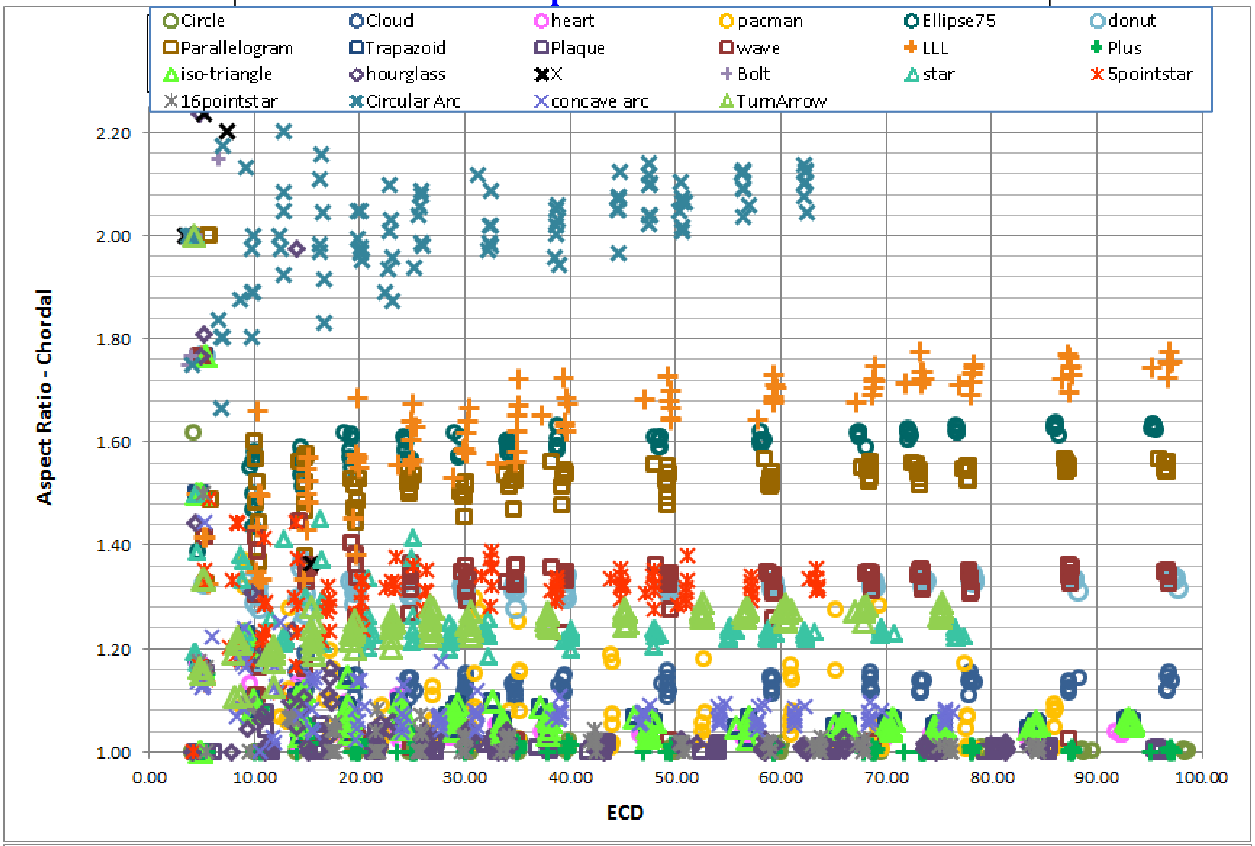
Bottom: Same plot as above, but for shapes with aspect ratio > 10 (Ellipse 5 & 10) and Aspect ratio 3-10 (Ellipse 20 & 30). These uncertainties need to be confirmed with a set of convex and non-ellipse shapes.

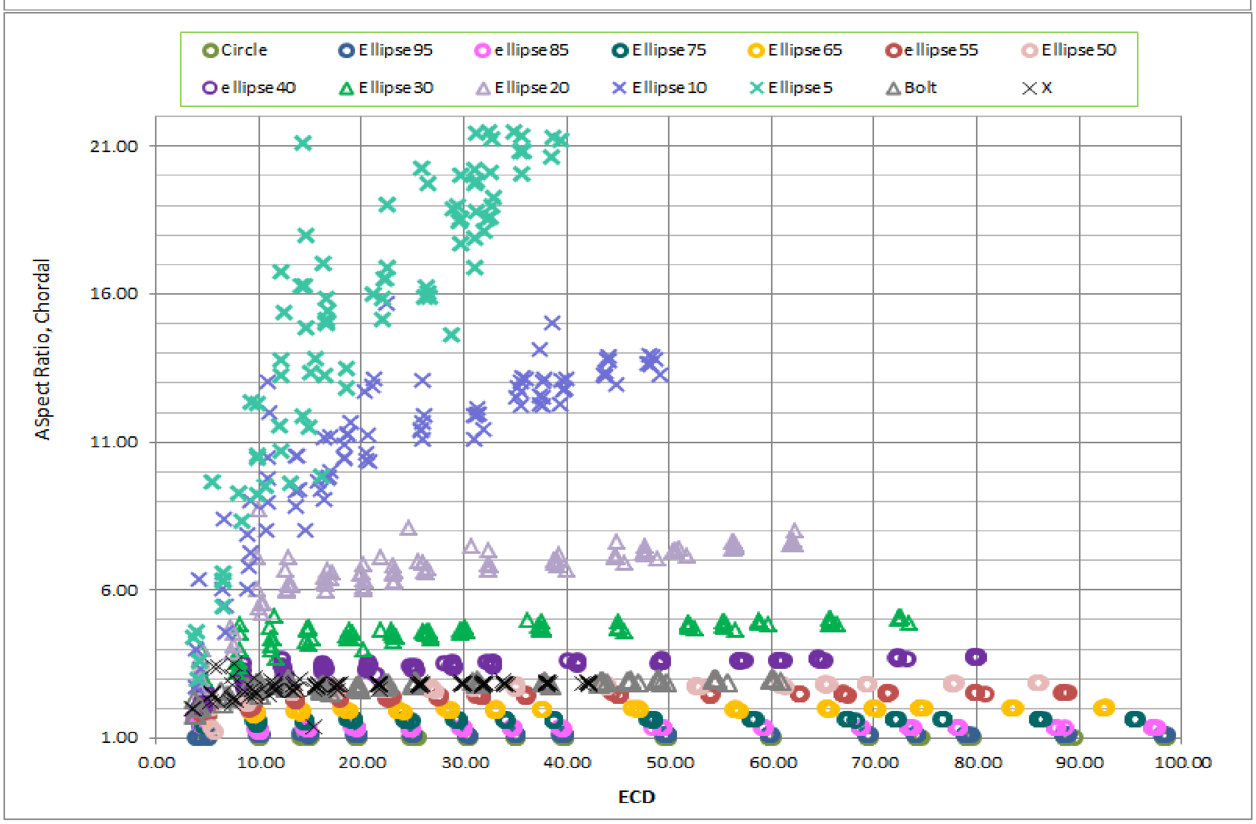
Raw Data Plots: Aspect ratio Ellipse versus ECD.



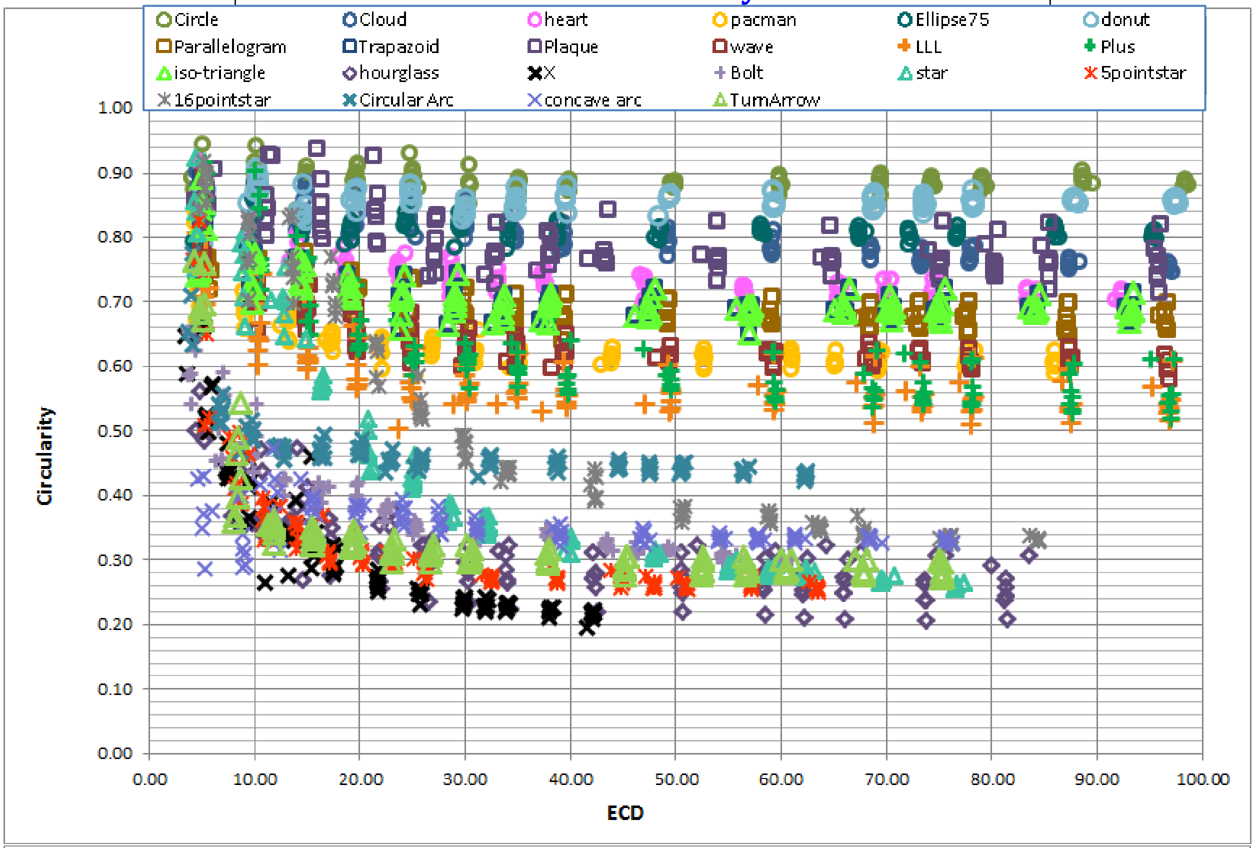


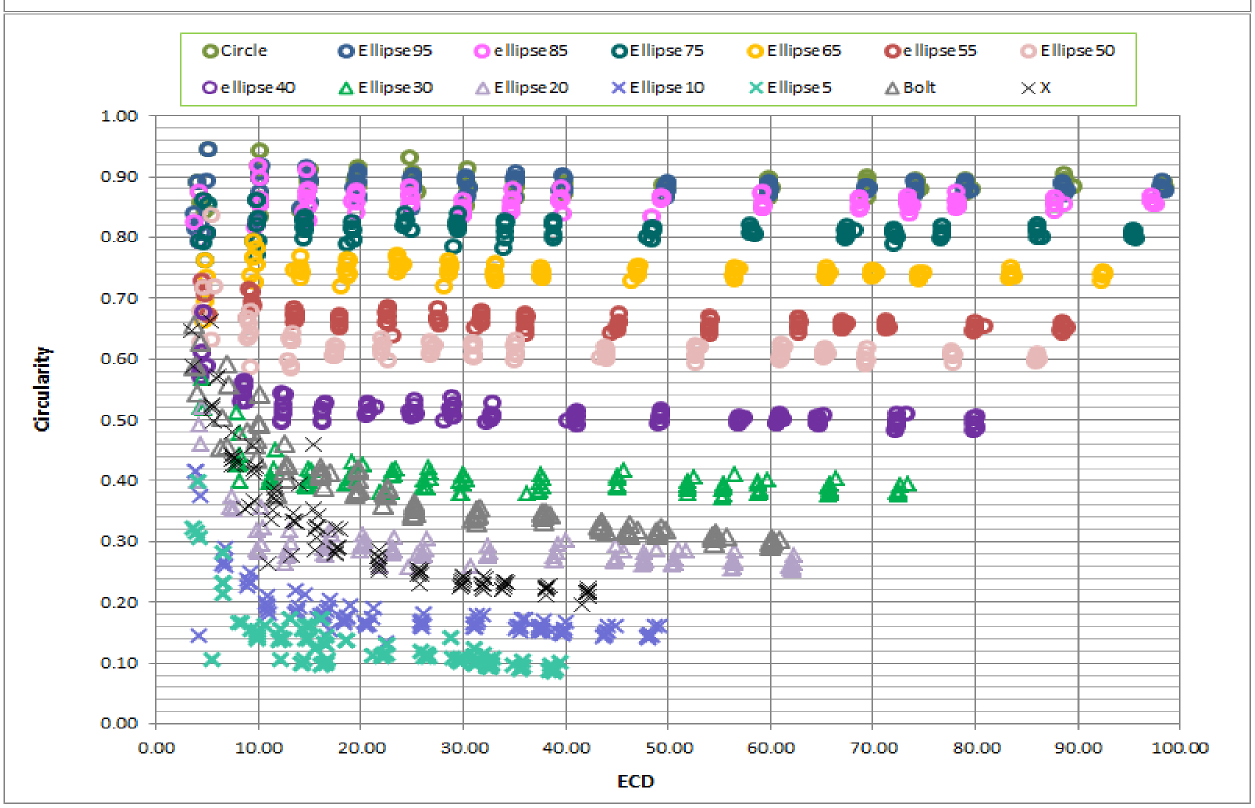
Raw Data Plots: Aspect ratio Chordal versus ECD.



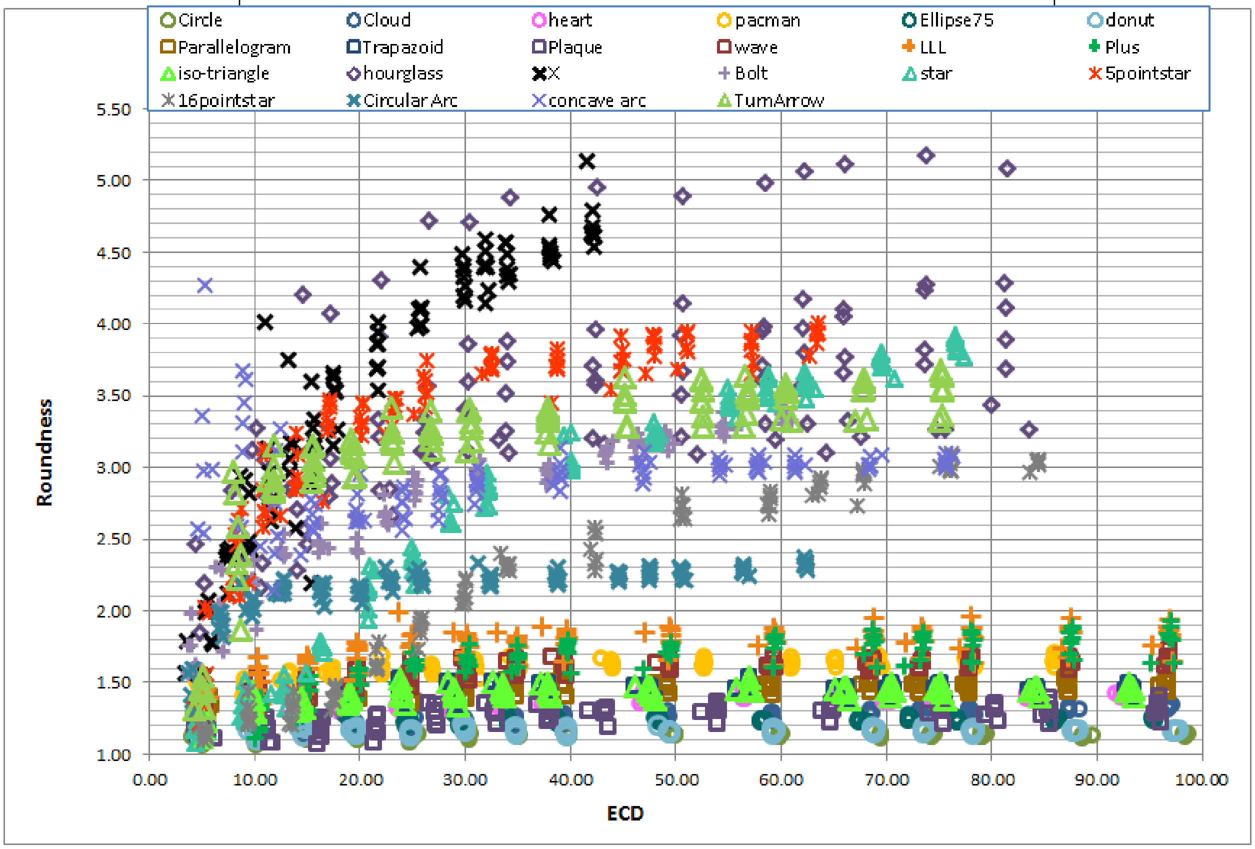


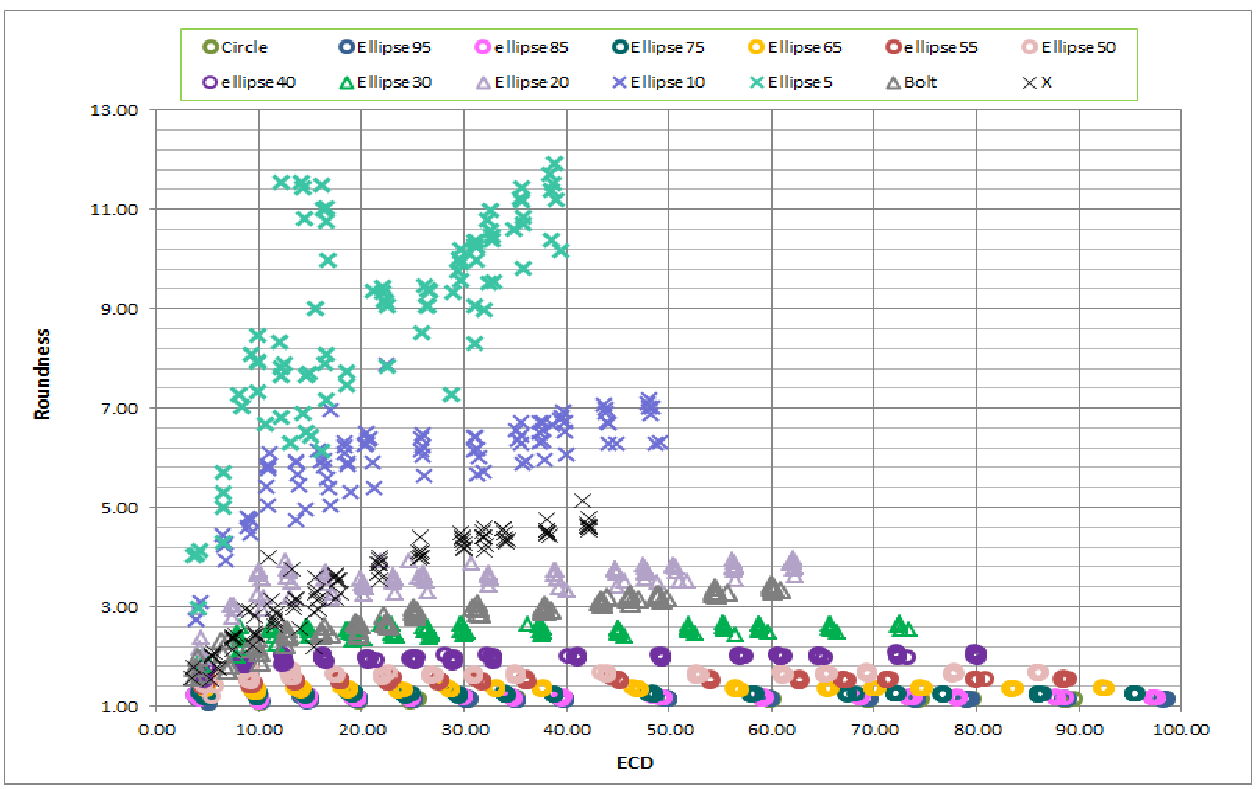
Raw Data Plots: Circularity versus ECD.



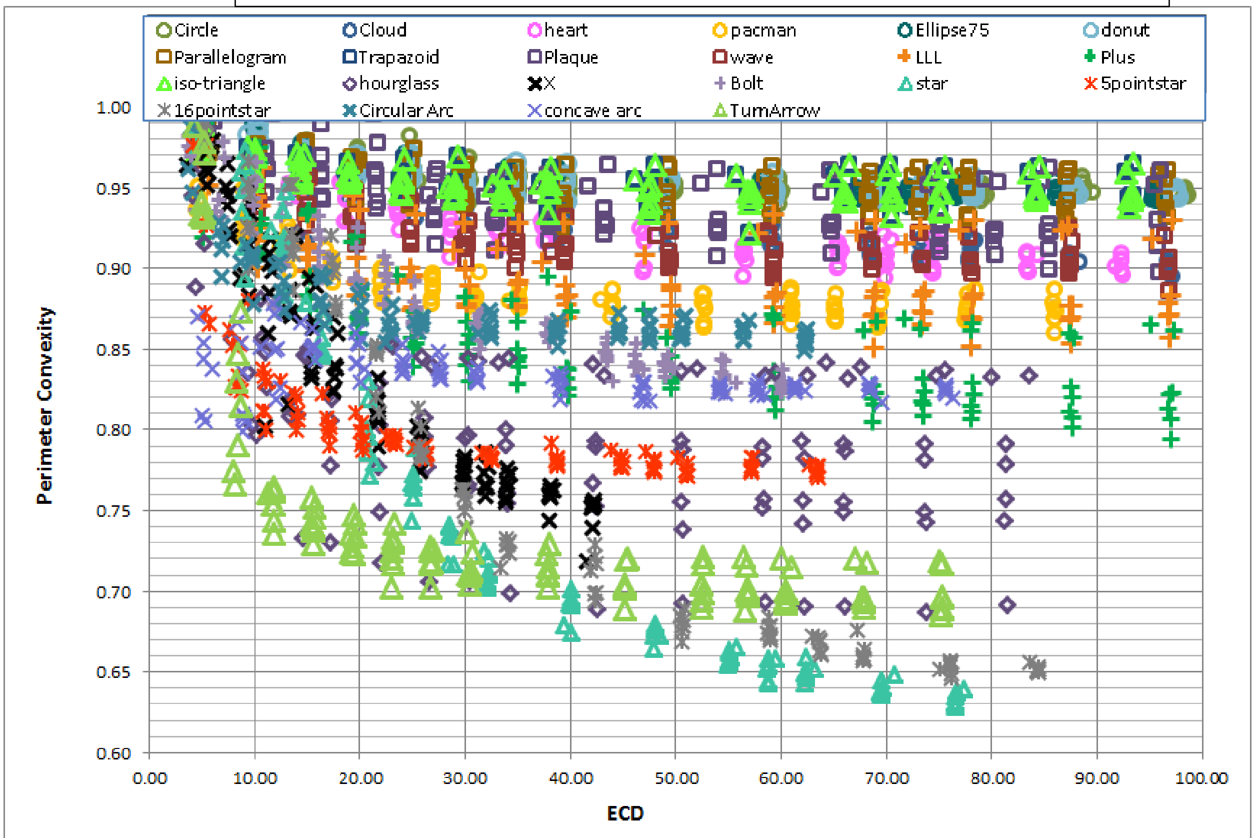


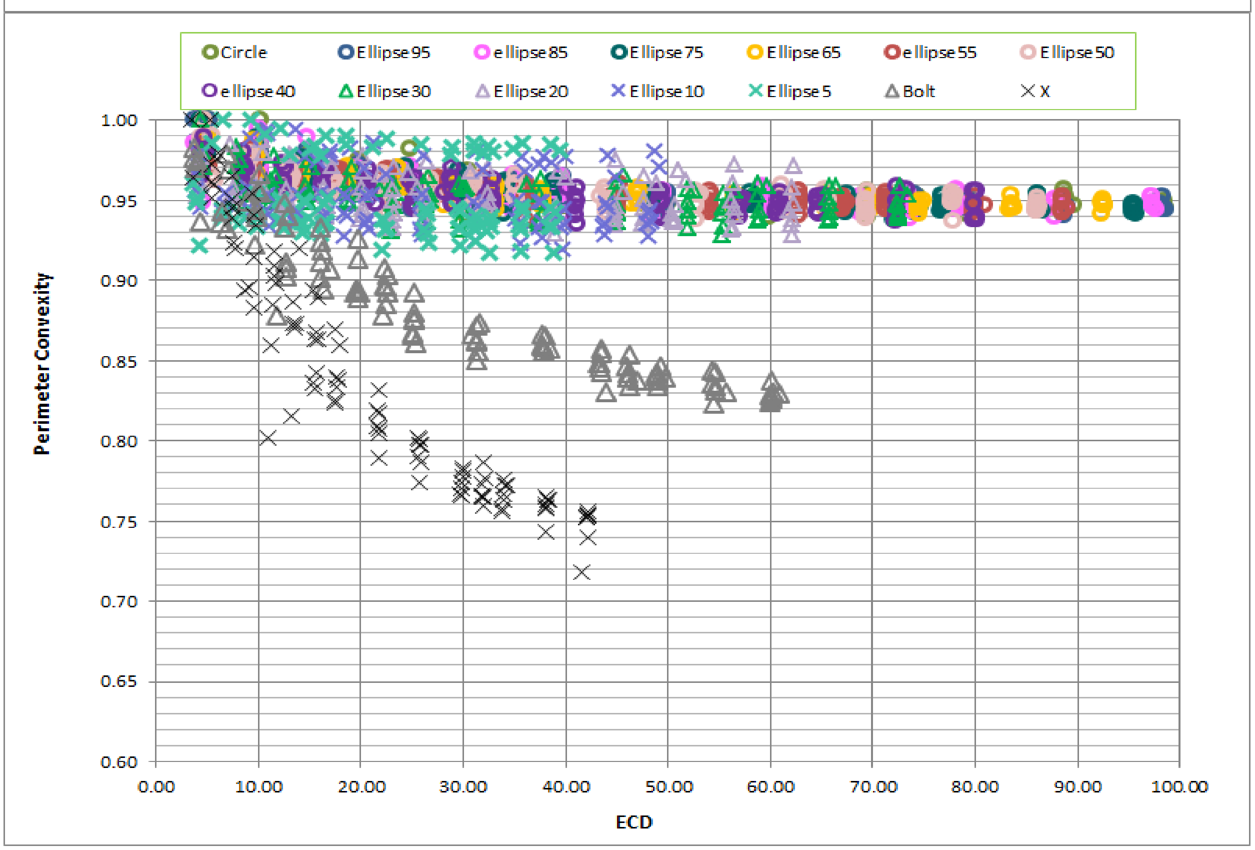
Raw Data Plots: Roundness versus ECD.



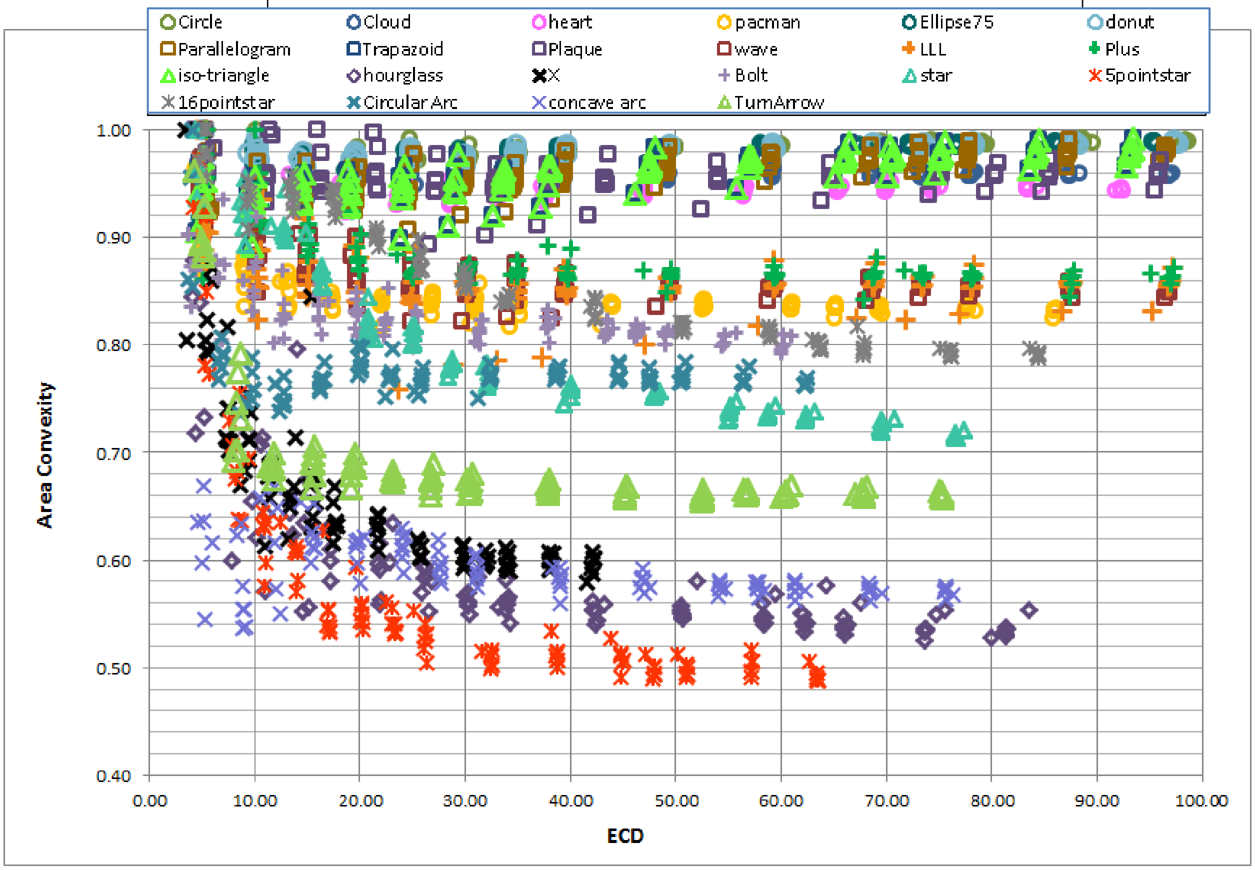


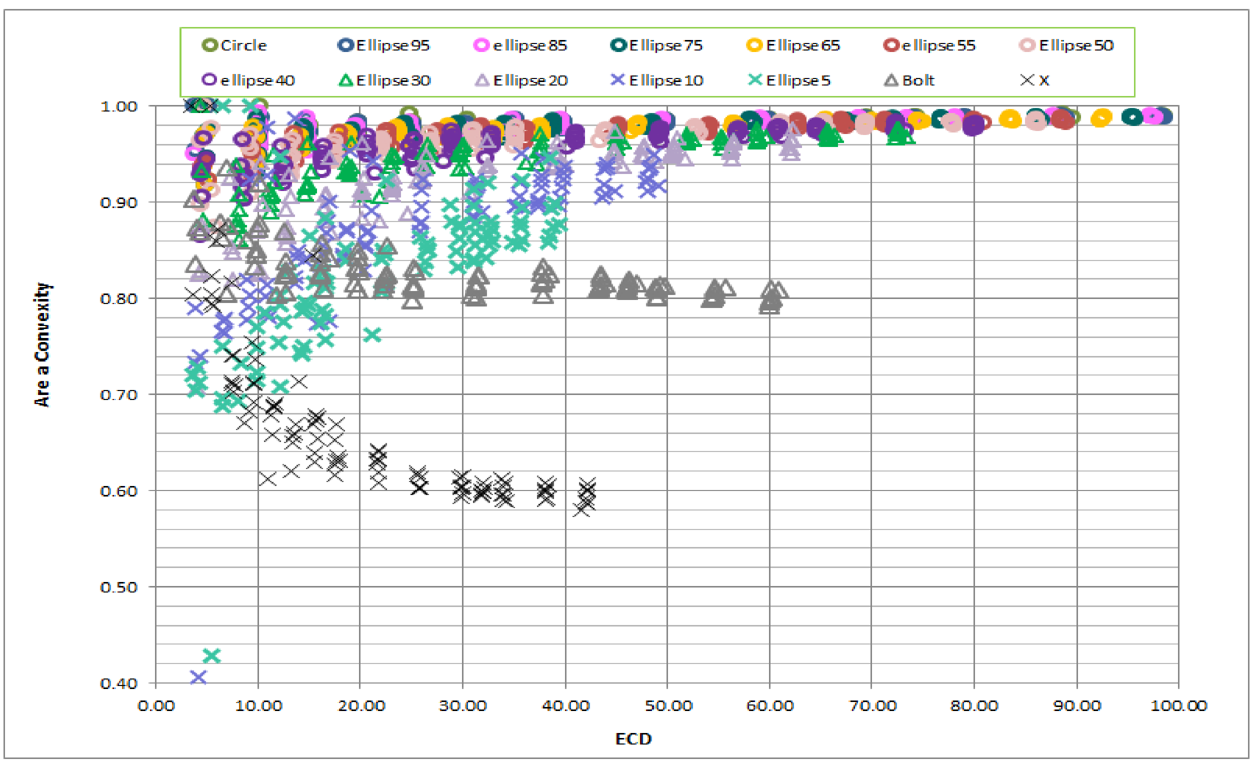
Raw Data Plots: Perimeter Convexity versus ECD.



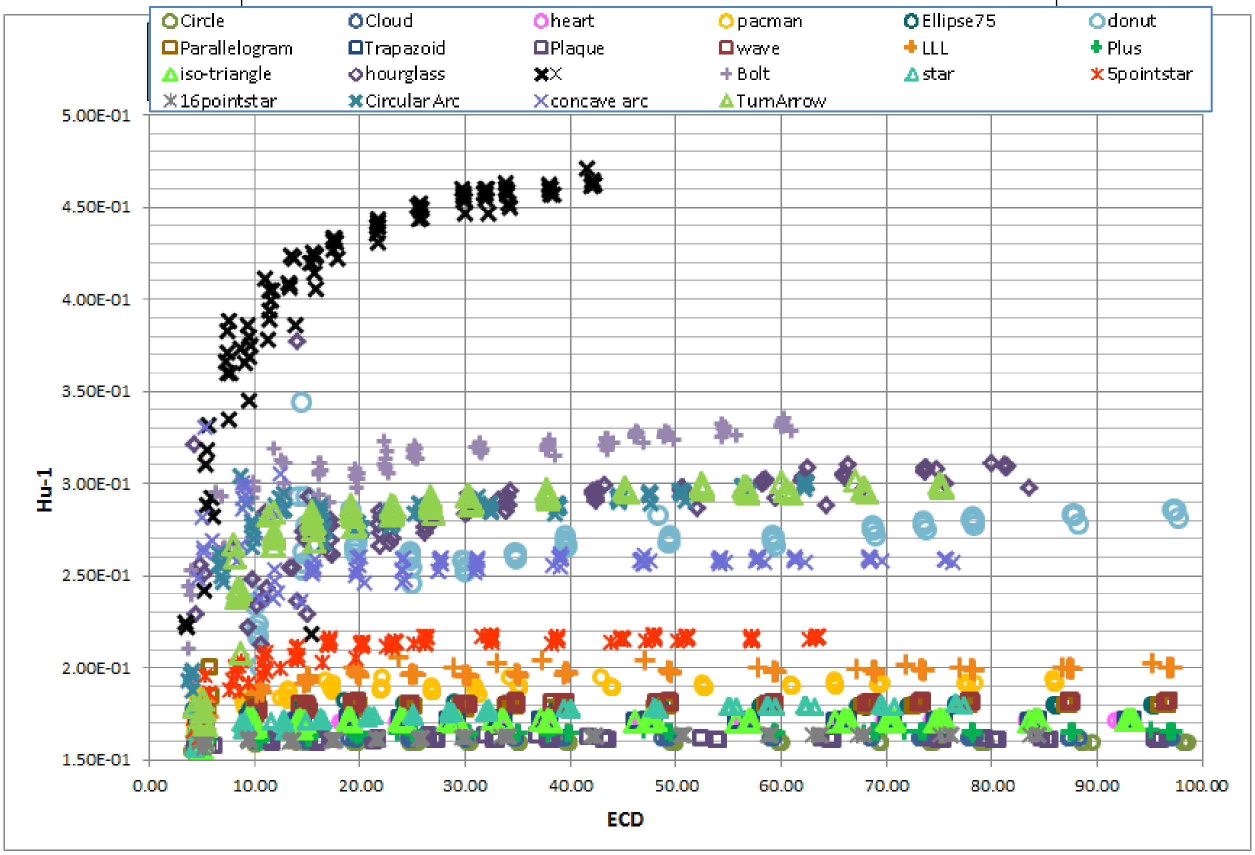


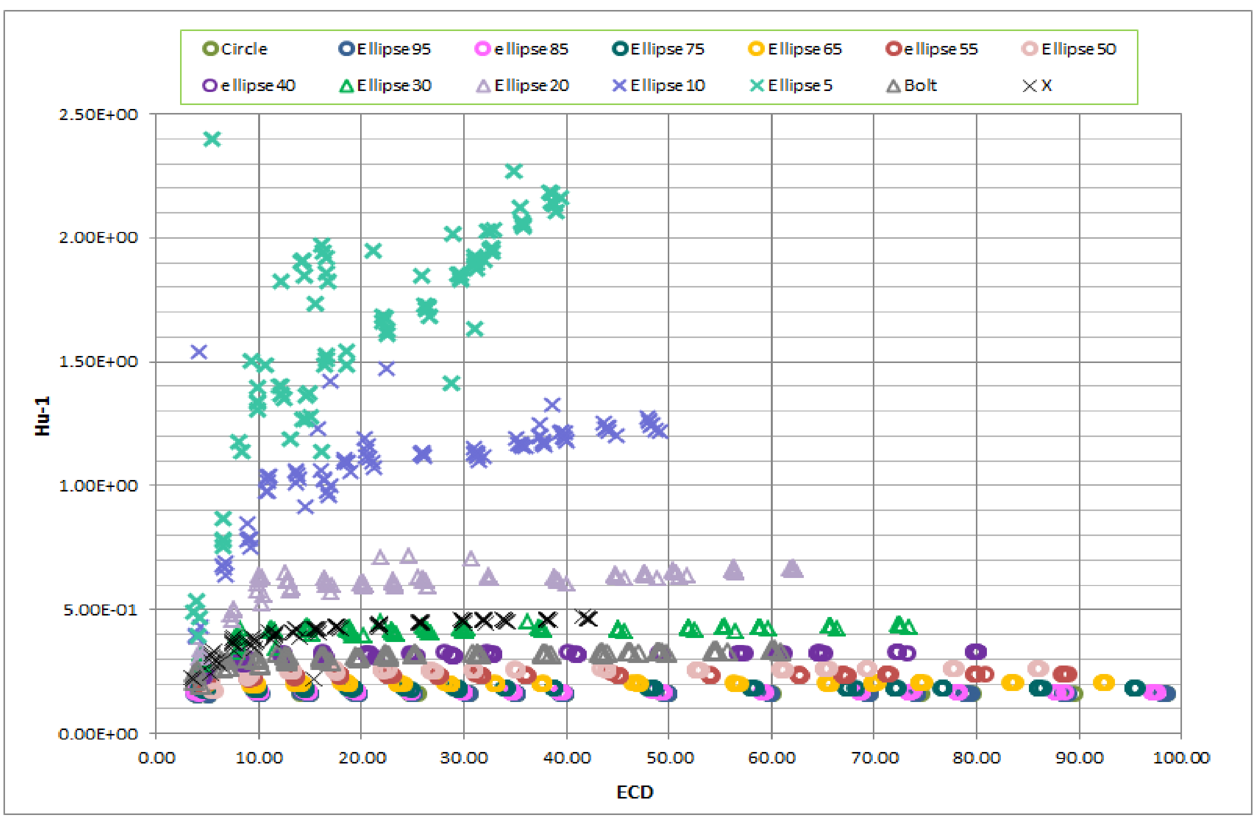
Raw Data Plots: Area Convexity versus ECD.





Raw Data Plots: Hu-1 versus ECD.





Raw Data Plots: Hu-2 versus ECD.